

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 14, 2003, 04:14:55 ; Search time 111 Seconds  
(without alignments)  
5765.816 Million cell updates/sec

Title: US-09-701-229-1

Perfect score: 1450

Sequence: 1 cgtgctgctgctgctgcca.....tcttgagcggcagcgcacgc 1450

Scoring table: OLIGO.NUC

Gapop 60.0 , Gapext 60.0

Searched: 569978 seqs, 220691566 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

Issued\_Patents\_NA.\*

- 1: /cgn2.6/ptodata/1/ina/5A\_COMB.seq.\*
- 2: /cgn2.6/ptodata/1/ina/5B\_COMB.seq.\*
- 3: /cgn2.6/ptodata/1/ina/6A\_COMB.seq.\*
- 4: /cgn2.6/ptodata/1/ina/6B\_COMB.seq.\*
- 5: /cgn2.6/ptodata/1/ina/PCTUS\_COMB.seq.\*
- 6: /cgn2.6/ptodata/1/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1183	81.6	1401	4	US-09-252-991A-7861
2	1167	80.5	1371	4	US-09-252-991A-7702
3	567	39.1	567	4	US-09-252-991A-7787
4	438	30.2	564	4	US-09-252-991A-7928
5	258	17.8	1359	4	US-09-252-991A-7624
6	160	11.0	360	4	US-09-252-991A-7552
7	120	8.3	1170	4	US-09-252-991A-7929
8	62	4.3	1287	4	US-09-252-991A-7860
9	36	2.5	1404	4	US-09-252-991A-7623
10	19	1.3	204	4	US-09-252-991A-9277
11	19	1.3	405	4	US-09-252-991A-1743
12	19	1.3	414	4	US-09-252-991A-12452
13	19	1.3	489	4	US-09-252-991A-8304
14	19	1.3	522	4	US-09-072-596-263
15	19	1.3	723	4	US-09-252-991A-13195
16	19	1.3	873	4	US-09-252-991A-4459
17	19	1.3	888	4	US-09-252-991A-4147
18	19	1.3	930	4	US-09-252-991A-8815
19	19	1.3	942	4	US-09-252-991A-12760
20	19	1.3	966	4	US-09-252-991A-4666
21	19	1.3	1116	4	US-09-252-991A-10803
22	19	1.3	1206	4	US-09-252-991A-10627
23	19	1.3	1275	4	US-09-252-991A-10722
24	19	1.3	1641	4	US-09-252-991A-2066
25	19	1.3	1764	4	US-09-252-991A-16183
26	19	1.3	1770	4	US-09-252-991A-16408
27	19	1.3	1806	4	US-09-252-991A-15978

c	28	19	1.3	2112	4	US-09-252-991A-1820	Sequence 1820, Ap
	29	19	1.3	2283	4	US-09-252-991A-2140	Sequence 2140, Ap
	30	19	1.3	2932	4	US-09-016-434-1419	Sequence 1419, Ap
	31	19	1.3	4143	4	US-09-252-991A-9151	Sequence 9151, Ap
c	32	19	1.3	4212	4	US-09-252-991A-8929	Sequence 8929, Ap
c	33	19	1.3	14272	4	US-09-516-914-23	Sequence 23, Appl
	34	19	1.3	4403765	3	US-09-103-840A-2	Sequence 2, Appl
	35	19	1.3	4411529	3	US-09-103-840A-1	Sequence 1, Appl
	36	18	1.2	603	4	US-09-252-991A-1076	Sequence 1076, Ap
	37	18	1.2	792	4	US-09-252-991A-7005	Sequence 7005, Ap
	38	18	1.2	1308	4	US-09-501-115-31	Sequence 31, Appl
	39	18	1.2	1383	4	US-09-252-991A-16519	Sequence 16519, A
c	40	18	1.2	1488	4	US-09-252-991A-5798	Sequence 5798, Ap
c	41	18	1.2	1683	4	US-09-252-991A-10216	Sequence 10216, A
c	42	18	1.2	1818	4	US-09-252-991A-9837	Sequence 9837, Ap
	43	18	1.2	1842	4	US-09-252-991A-5853	Sequence 5853, Ap
	44	18	1.2	1893	4	US-09-252-991A-9753	Sequence 9753, Ap
c	45	18	1.2	2028	4	US-09-252-991A-7091	Sequence 7091, Ap

ALIGNMENTS

RESULT 1

US-09-252-991A-7861/c  
; Sequence 7861, Application US/09252991A.  
; Patent No. 6551795

GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

; FILE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: 107196.136

; CURRENT APPLICATION NUMBER: US/09/252,991A

; CURRENT FILING DATE: 1999-02-18

; PRIOR APPLICATION NUMBER: US 60/074,788

; PRIOR FILING DATE: 1998-02-18

; PRIOR APPLICATION NUMBER: US 60/094,190

; PRIOR FILING DATE: 1998-07-27

; NUMBER OF SEQ ID NOS: 33142

; SEQ ID NO 7861

; LENGTH: 1401

; TYPE: DNA

; ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-7861

Query Match 81.6%; Score 1183; DB 4; Length 1401;  
Best Local Similarity 99.7%; Pred.No. 0;  
Matches 1383; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY	2	GTGCTGATCGGCTCGCCACCTTGAAGCTGCGTTGAGGACGAAGAGAGATGAGCTGAT	61
DB	1387	GTGCTGATCGGCTCGCCACCTTGAAGCTGCGTTGAGGACGAAGAGAGATGAGCTGAT	1328
QY	62	CGCCTCGGACCACTTCCGCGATCGTTCGGCCCTCGGCAAGAGCGGCGATGCTCTGCTGG	121
DB	1327	CGCCTCGGACCACTTCCGCGATCGTTCGGCCCTCGGCAAGAGCGGCGATGCTCTGCTGG	1268
QY	122	CTACCTGGCGCGCGCGCTTCCCTTTCGCGGTCGATACCGAGAGAACCCCGCGGA	181
DB	1267	CTACCTGGCGCGCGCGCTTCCCTTTCGCGGTCGATACCGAGAGAACCCCGCGGA	1208
QY	182	GCTGGCCACCTTCCGCGAGTATCCGAGGTGGAAGTTCGTTGGCGGAACCTCGACGC	241
DB	1207	GCTGGCCACCTTCCGCGAGTATCCGAGGTGGAAGTTCGTTGGCGGAACCTCGACGC	1148
QY	242	CGAGTTCCTCTGCTCGGCGCGGAACTATGTCAGCCCGGCTTGTCTCGCGGACCC	301
DB	1147	CGAGTTCCTCTGCTCGGCGCGGAACTATGTCAGCCCGGCTTGTCTCGCGGACCC	1088
QY	302	TCCGCTGTACAGCGCGCGGAAAGGCGGCGATCTCCGTCGATCATGATCTCTTCGC	361
DB	1087	TCCGCTGTACAGCGCGCGGAAAGGCGGCGATCTCCGTCGATCATGATCTCTTCGC	1028

QY 362 CCGCGAGCGAAGCCCGCGATCGTGGCCATCACCGTTCCAAACGCGAAGACACCGTGAC 421  
 Db 1027 CCGCGAGCGAAGCCCGCGATCGTGGCCATCACCGTTCCAAACGCGAAGACACCGTGAC 968  
 QY 422 CACCTGTGTGGCGAATGGCGGTGGCGGAGAACGGTGTCCGCTCGCGCGCAACCT 481  
 Db 967 CACCTGTGTGGCGAATGGCGGTGGCGGAGAACGGTGTCCGCTCGCGCGCAACCT 908  
 QY 482 CGGACACCCCGCGCTCGACCTGCTGGCGGAGACATCGAGCTGTACGTGTGGAGCTGTC 541  
 Db 907 CGGACACCCCGCGCTCGACCTGCTGGCGGAGACATCGAGCTGTACGTGTGGAGCTGTC 848  
 QY 542 GAGCTTCAGCTGGAACCTGCGATCGCTCAACGCGAGGTGGCGACCGTGTGAAAGT 601  
 Db 847 GAGCTTCAGCTGGAACCTGCGATCGCTCAACGCGAGGTGGCGACCGTGTGAAAGT 788  
 QY 602 CAGCGAAGACCATATGATCGCTACGAGCGCATGGCTGACTACCACTGGCCCAAGCACCG 661  
 Db 787 CAGCGAAGACCATATGATCGCTACGAGCGCATGGCTGACTACCACTGGCCCAAGCACCG 728  
 QY 662 GATCTTCGGGGTCCCGCCAGGTGCTGGTGAATCGCGCGATGCCCTGACCCGACCGCT 721  
 Db 727 GATCTTCGGGGTCCCGCCAGGTGCTGGTGAATCGCGCGATGCCCTGACCCGACCGCT 668  
 QY 722 GATCGCGATACCGTGGCTGCTGGTTCGGCTGAACAAGCCGACCTTCAAGGCTTT 781  
 Db 667 GATCGCGATACCGTGGCTGCTGGTTCGGCTGAACAAGCCGACCTTCAAGGCTTT 608  
 QY 782 CGGCTGATCGAGGAAGACGCGCAAGTGGCTGGCTTCAGTTCGACAAAGCTGTGCC 841  
 Db 607 CGGCTGATCGAGGAAGACGCGCAAGTGGCTGGCTTCAGTTCGACAAAGCTGTGCC 548  
 QY 842 GGTGGCGAATGAAGTCCGTGGCGCCCAACATATTCGAACGCGTGGCGGCGTGGC 901  
 Db 547 GGTGGCGAATGAAGTCCGTGGCGCCCAACATATTCGAACGCGTGGCGGCGTGGC 488  
 QY 902 GCTGGGCGATCGGCTGGCTGCGTTCGAGCGCATGCTGGCGCGTGAAGGCGTTTC 961  
 Db 487 GCTGGGCGATCGGCTGGCTGCGTTCGAGCGCATGCTGGCGCGTGAAGGCGTTTC 428  
 QY 962 CGGCTGGCTCATCGCTGCGAGTGGGTACGCGAGCGCGAGGGCTGAGCTACTACGACA 1021  
 Db 427 CGGCTGGCTCATCGCTGCGAGTGGGTACGCGAGCGCGAGGGCTGAGCTACTACGACA 368  
 QY 1022 TTCAAAGGCCAACCACTCGGCGCGCCCTGGCGCGATCGAGGGGCTGGGTGCCGACAT 1081  
 Db 367 TTCAAAGGCCAACCACTCGGCGCGCCCTGGCGCGATCGAGGGGCTGGGTGCCGACAT 308  
 QY 1082 CGAGGCGAAGCTGGTGTGCTCGCGCGCGAGAGCGCAAGGGCGCCGATTTCCATGACCT 1141  
 Db 307 CGAGGCGAAGCTGGTGTGCTCGCGCGCGAGAGCGCAAGGGCGCCGATTTCCATGACCT 248  
 QY 1142 CGCGAGCGGCTCGCGGCTTCTCGCGCGCGGTGTACTGCTGGCGGTGACGCGCGGCT 1201  
 Db 247 CGCGAGCGGCTCGCGGCTTCTCGCGCGCGGTGTACTGCTGGCGGTGACGCGCGGCT 188  
 QY 1202 GATTGCCAGGCACTGGGCAACCGGTACCGCTGGTGGCGGTGCGCAACGCTGGACGAAGC 1261  
 Db 187 GATTGCCAGGCACTGGGCAACCGGTACCGCTGGTGGCGGTGCGCAACGCTGGACGAAGC 128  
 QY 1262 AGTCCGCGAGCGCGCGAGCTGGCGCGCGAGGGGATGCGGTGCTGTTCGCGCGGCTG 1321  
 Db 127 AGTCCGCGAGCGCGCGAGCTGGCGCGCGAGGGGATGCGGTGCTGTTCGCGCGGCTG 68  
 QY 1322 CGGAGCGCTGGACATGTTCAAGAACTTCGAAGAACGCGGAGCGGCTGTTCCGCAAGCGCT 1381  
 Db 67 CGGAGCGCTGGACATGTTCAAGAACTTCGAAGAACGCGGAGCGGCTGTTCCGCAAGCGCT 8  
 QY 1382 AGAGGAG 1388  
 Db 7 AGAGGAG 1

RESULT 2  
 US-09-252-991A-7702  
 ; Sequence 7702, Application US/09252991A  
 ; Patent No. 6551795  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Marc J. Rubenfield et al.  
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
 ; FILE REFERENCE: 107196.136  
 ; CURRENT APPLICATION NUMBER: US/09/252.991A  
 ; CURRENT FILING DATE: 1999-02-18  
 ; PRIOR APPLICATION NUMBER: US 60/074,788  
 ; PRIOR FILING DATE: 1998-02-18  
 ; PRIOR APPLICATION NUMBER: US 60/094,190  
 ; PRIOR FILING DATE: 1998-07-27  
 ; NUMBER OF SEQ ID NOS: 33142  
 ; SEQ ID NO 7702  
 ; LENGTH: 1371  
 ; TYPE: DNA  
 ; ORGANISM: Pseudomonas aeruginosa  
 US-09-252-991A-7702

Query Match 80.5%; Score 1167; DB 4; Length 1371;  
 Best Local Similarity 99.7%; Pred. No. 0;  
 Matches 1367; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 27 AGCTGCGCTTGAGGACGAAGAGAGATGAGCTGATCGCTCCGACCACTTCCGCATCGTT 86  
 Db 1 AGCTGCGCTTGAGGACGAAGAGAGATGAGCTGATCGCTCCGACCACTTCCGCATCGTT 60  
 QY 87 GTGCGCTCGGCAAGAGCGCATGTCCTGTGTGGCTACCTTGGCGCGCGCGGCTTGCCT 146  
 Db 61 GTGCGCTCGGCAAGAGCGCATGTCCTGTGTGGCTACCTTGGCGCGCGCGGCTTGCCT 120  
 QY 147 TTGCGCGTGTGCTATACCCGAGAACCCCGCGAGCTGGCCACCTTGGTGGCCAGTAT 206  
 Db 121 TTGCGCGTGTGCTATACCCGAGAACCCCGCGAGCTGGCCACCTTGGTGGCCAGTAT 180  
 QY 207 CCGCAGGTGGAAGTGGCTTGGCGGAACTCGACCCGAGTTCCTCTCTCGCCCGCGAA 266  
 Db 181 CCGCAGGTGGAAGTGGCTTGGCGGAACTCGACCCGAGTTCCTCTCTCGCCCGCGAA 240  
 QY 267 CTCTATGTACGCCCGCGCTTGTGCTGCGCACCCCTGCGGTGTGACAGGCGCGCGAAA 326  
 Db 241 CTCTATGTACGCCCGCGCTTGTGCTGCGCACCCCGCGGTGTGACAGGCGCGCGAAA 300  
 QY 327 GCGGTGCGCATCTCCGCTGACATGATCTCTTGGCGCGAGGCGAAGGCCCGATGTC 386  
 Db 301 GGTGTGCGCATCTCCGCTGACATGATCTCTTGGCGCGAGGCGAAGGCCCGATGTC 360  
 QY 387 GCCATCACCGGTTCCAAACGGAAGACACCGTGAACACCTGTGGCGGAAATGGCGGTG 446  
 Db 361 GCCATCACCGGTTCCAAACGGAAGACACCGTGAACACCTGTGGCGGAAATGGCGGTG 420  
 QY 447 GCGCGGACAAAGCGTGTGCGCGTGGCGGCAACCTCGCGACCCCGCGGCTGACGCTGTC 506  
 Db 421 GCGCGGACAAAGCGTGTGCGCGTGGCGGCAACCTCGCGACCCCGCGGCTGACGCTGTC 480  
 QY 507 GCGCGACATCGAGCTGTACGTGTGGAGCTGTTCAGCTTCCAGCTTGAACCTGCGAT 566  
 Db 481 GCGCGACATCGAGCTGTACGTGTGGAGCTGTTCAGCTTCCAGCTTGAACCTGCGAT 540  
 QY 567 CGGCTCAACCGCGAGGTGGCGACCGTGTGAAGTTCAGCGAAGACCAATATGGATCGTAC 626  
 Db 541 CGGCTCAACCGCGAGGTGGCGACCGTGTGAAGTTCAGCGAAGACCAATATGGATCGTAC 600  
 QY 627 GAGCGATGCTGACTACCACTGGCGCAACGACCGGATCTTCGCGGTGCGCGCGAGTTC 686  
 Db 601 GAGCGATGCTGACTACCACTGGCGCAACGACCGGATCTTCGCGGTGCGCGCGAGTTC 660  
 QY 687 GTGTTGAATCGCGCGGATGCCCTGACCCGACCGTGTATCGCCGATACCGTGGCTGCTGG 746  
 Db 661 GTGTTGAATCGCGCGGATGCCCTGACCCGACCGTGTATCGCCGATACCGTGGCTGCTGG 720

747 TCCTGCGCTGAACAGCGGACTTCAAGGCTTTTCGGCTGATCGAGGAGACGGCCAG 806  
 721 TCCTGCGCTGAACAGCGGACTTCAAGGCTTTTCGGCTGATCGAGGAGACGGCCAG 780  
 807 AAGTGGCTGGGTTCCAGTTCGACAAAGCTCTGCGCGGTTGGCAACTGAAGATCCGTTGC 866  
 781 AAGTGGCTGGGTTCCAGTTCGACAAAGCTCTGCGCGGTTGGCAACTGAAGATCCGTTGC 840  
 867 GCCACAACTATTCACAGCGCTCGCGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCG 926  
 841 GCCACAACTATTCACAGCGCTCGCGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCG 900  
 927 TTCAGACCCATGCTCGCGCGCTGAAGCGCTTTTCGGCGCTGCGCTGCGCTGCGCTGCG 985  
 901 TTCAGACCCATGCTCGCGCGCTGAAGCGCTTTTCGGCGCTGCGCTGCGCTGCGCTGCG 960  
 987 GTACGAGCGGCGAGGCGTGAAGCTACTACGACGATTCACAGGCGACCAACGTCGCGGCC 1046  
 961 GTACGAGCGGCGAGGCGTGAAGCTACTACGACGATTCACAGGCGACCAACGTCGCGGCC 1020  
 1047 GCCCTGCGCGATCGAGGCGTGGGTGCGACATCGACGAGCTGGTGGTGGTGGTGGTGG 1106  
 1021 GCCCTGCGCGATCGAGGCGTGGGTGCGACATCGACGAGCTGGTGGTGGTGGTGGTGG 1080  
 1107 GCGGAGACGCGAAGGCGCGGATTTCCATGACCTGCGCGAGCGGTCGCGCGCTTCTGC 1166  
 1081 GCGGAGACGCGAAGGCGCGGATTTCCATGACCTGCGCGAGCGGTCGCGCGCTTCTGC 1140  
 1167 CGGCGGTGTACTGCTTGGCGGTGACGCGCGGCTGATTCGCCAGGCACTGGGCAAGCG 1226  
 1141 CGGCGGTGTACTGCTTGGCGGTGACGCGCGGCTGATTCGCCAGGCACTGGGCAAGCG 1200  
 1227 GTACCGCTGTGCGCTCGCAACCTGGAAGCACTGCGGAGCGCGCGGCTGCGGCGGCG 1286  
 1201 GTACCGCTGTGCGCTCGCAACCTGGAAGCACTGCGGAGCACTGCGGAGCGCGGCGG 1260  
 1287 CGCAAGCGATGCGGTGCTGTTGTCGCCGCGCTGCGGAGCGCTGAGATGTTCAAGAAC 1346  
 1261 CGCAAGCGATGCGGTGCTGTTGTCGCCGCGCTGCGGAGCGCTGAGATGTTCAAGAAC 1320  
 1347 TTCAGAGACGCGACGCGCTTTCGCCAAGCGCTAGAGGAGCTAGCGTGA 1397  
 1321 TTCAGAGACGCGACGCGCTTTCGCCAAGCGCTAGAGGAGCTAGCGTGA 1371

RESULT 3  
 US-09-252-991A-7787/c  
 ; Sequence 7787, Application US/09252991A  
 ; Patent No. 6551795  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Marc J. Rubenfield et al.  
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
 ; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
 ; FILE REFERENCE: 107196.136  
 ; CURRENT APPLICATION NUMBER: US/09/252.991A  
 ; CURRENT FILING DATE: 1999-02-18  
 ; PRIOR APPLICATION NUMBER: US 60/074,788  
 ; PRIOR FILING DATE: 1998-02-18  
 ; PRIOR APPLICATION NUMBER: US 60/094,190  
 ; PRIOR FILING DATE: 1998-07-27  
 ; NUMBER OF SEQ ID NOS: 33142  
 ; SEQ ID NO 7787  
 ; LENGTH: 567  
 ; TYPE: DNA  
 ; ORGANISM: Pseudomonas aeruginosa  
 ; US-09-252-991A-7787

Query Match 39.1%; Score 567; DB 4; Length 567;  
 Best Local Similarity 100.0%; Pred. No. 1e-251;  
 Matches 567; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

772 TCAAGGCTTTTCGGCTGATCGAGGAGACGGCCAGAGTGGCTGGCGTTCCAGTTCGACA 831

567 TCAAGGCTTTTCGGCTGATCGAGGAGACGGCCAGAGTGGCTGGCGTTCCAGTTCGACA 508  
 832 AGCTGCTGCGGTTGGCGAACTGAAGATCCGTGGCGGCCCAACAATATTTCAACGCGCTCG 891  
 507 AGCTGCTGCGGTTGGCGAACTGAAGATCCGTGGCGGCCCAACAATATTTCAACGCGCTCG 448  
 892 CCGCGCTGGGCGTGGGCGATCGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCG 951  
 447 CCGCGCTGGGCGTGGGCGATCGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCG 388  
 952 AGCGTTTTCGGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCT 1011  
 387 AGCGTTTTCGGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCT 328  
 1012 ACTACGACGATTCACAGGCGCACCAACGTCGCGCGCGCGCGCTGCGCGCGCTGCGCGCG 1071  
 327 ACTACGACGATTCACAGGCGCACCAACGTCGCGCGCGCGCGCTGCGCGCGCTGCGCGCG 268  
 1072 GTGCCGACATCGAGCGCAAGCTGGTGTGCTGCTGCGCGCGCGAGAGCGCAAGGCGCGGATT 1131  
 267 GTGCCGACATCGAGCGCAAGCTGGTGTGCTGCTGCGCGCGCGAGAGCGCAAGGCGCGGATT 208  
 1132 TCATGACCTGCGGAGCGCGCTGCGCGCTGCTGCGCGCGCGAGAGCGCAAGGCGCGGATT 1191  
 207 TCCATGACCTGCGGAGCGCGCTGCGCGCTTCTGCGCGCGCGGCTGCTGCTGCGCGCTG 148  
 1192 ACGCCGCGCTGATTTGCCAGGCACTGGGCAACGCGGTACCGCTGGTGGCTGCGCAACGC 1251  
 147 ACGCCGCGCTGATTTGCCAGGCACTGGGCAACGCGGTACCGCTGGTGGCTGCGCAACGC 88  
 1252 TGGACGAAGCAGTCCGCGAGCGCGCGAGCTGGCGCGCGAGAGCGATCGCGTGTGTTGT 1311  
 87 TGGACGAAGCAGTCCGCGAGCGCGCGAGCTGGCGCGCGAGAGCGATCGCGTGTGTTGT 28  
 1312 CGCGGCGCTGCGGAGCGCTGGACATGT 1338  
 27 CGCGGCGCTGCGGAGCGCTGGACATGT 1

RESULT 4  
 US-09-252-991A-7928/c  
 ; Sequence 7928, Application US/09252991A  
 ; Patent No. 6551795  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Marc J. Rubenfield et al.  
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
 ; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
 ; FILE REFERENCE: 107196.136  
 ; CURRENT APPLICATION NUMBER: US/09/252.991A  
 ; CURRENT FILING DATE: 1999-02-18  
 ; PRIOR APPLICATION NUMBER: US 60/074,788  
 ; PRIOR FILING DATE: 1998-02-18  
 ; PRIOR APPLICATION NUMBER: US 60/094,190  
 ; PRIOR FILING DATE: 1998-07-27  
 ; NUMBER OF SEQ ID NOS: 33142  
 ; SEQ ID NO 7928  
 ; LENGTH: 564  
 ; TYPE: DNA  
 ; ORGANISM: Pseudomonas aeruginosa  
 ; US-09-252-991A-7928

Query Match 30.2%; Score 438; DB 4; Length 564;  
 Best Local Similarity 100.0%; Pred. No. 2.5e-192;  
 Matches 438; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1013 CTACGACGATTCACAGGCGCACCAACGTCGCGCGCGCGCGCTGCGCGCGATCGAGGCGCTGG 1072  
 564 CTACGACGATTCACAGGCGCACCAACGTCGCGCGCGCGCGCTGCGCGCGATCGAGGCGCTGG 505  
 1073 TGCGGACATCGAGCGCAAGCTGGTGTGCTGCTGCGCGCGGAGAGCGCAAGGCGCGCGATT 1132  
 504 TGCGGACATCGAGCGCAAGCTGGTGTGCTGCTGCGCGCGGAGAGCGCAAGGCGCGCGATT 445

```

QY 1133 CCATGACCTGCGGAGCGGTGCGCGCTTCTGCGCGGGTGGTACTGTTGGCGGTGA 1192
Db 444 CCATGACCTGCGGAGCGCGGTGCGCGCTTCTGCGCGGGTGGTACTGTTGGCGGTGA 385
QY 1193 CGCGGGCTGATTTGCCAGGCACTGGGCAACGCGGTACCGTGGTGGCGGTGCGCAACGCT 1252
Db 384 CGCGGGCTGATTTGCCAGGCACTGGGCAACGCGGTACCGTGGTGGCGGTGCGCAACGCT 325
QY 1253 GGACGAAGCAGTCCGCGAGCGCGCGAGCTGGCCCGGGAAGCGGATGCGGTGTTGTC 1312
Db 324 GGACGAAGCAGTCCGCGAGCGCGCGAGCTGGCCCGGGAAGCGGATGCGGTGTTGTC 265
QY 1313 GCGCGGCTGCGGAGCGCTGGACATGTTCAAGAACTTCGAAGAACGCGGACGCTGTTGCG 1372
Db 264 GCGCGGCTGCGGAGCGCTGGACATGTTCAAGAACTTCGAAGAACGCGGACGCTGTTGCG 205
QY 1373 CAAAGCCGTAGAGGAGCTAGCGTGTGTCGCGTGTGCGCCCTTCCCGTGGCGGTG 1432
Db 204 CAAAGCCGTAGAGGAGCTAGCGTGTGTCGCGTGTGCGCCCTTCCCGTGGCGGTG 145
QY 1433 TTGAGCGCGCACGGCATC 1450
Db 144 TTGAGCGCGCACGGCATC 127

```

RESULT 5  
 US-09-252-991A-7624  
 ; Sequence 7624, Application US/09252991A  
 ; Patent No. 6551795  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Marc J. Rubenfield et al.  
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
 ; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
 ; FILE REFERENCE: 107196.136  
 ; CURRENT APPLICATION NUMBER: US/09/252,991A  
 ; PRIOR FILING DATE: 1999-02-18  
 ; PRIOR APPLICATION NUMBER: US 60/074,788  
 ; PRIOR FILING DATE: 1998-02-18  
 ; PRIOR APPLICATION NUMBER: US 60/094,190  
 ; PRIOR FILING DATE: 1998-07-27  
 ; NUMBER OF SEQ ID NOS: 33142  
 ; SEQ ID NO 7624  
 ; LENGTH: 1359  
 ; TYPE: DNA  
 ; ORGANISM: Pseudomonas aeruginosa  
 US-09-252-991A-7624

```

Query Match 17.8%; Score 258; DB 4; Length 1359;
Best Local Similarity 100.0%; Pred. No. 1.6e-109;
Matches 258; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1193 CGCGGGCTGATTTGCCAGGCACTGGGCAACGCGGTACCGTGGTGGCGGTGCGCAACGCT 1252
Db 1 CGCGGGCTGATTTGCCAGGCACTGGGCAACGCGGTACCGTGGTGGCGGTGCGCAACGCT 60
QY 1253 GGACGAAGCAGTCCGCGAGCGCGCGAGCTGGCCCGGGAAGCGGATGCGGTGTTGTC 1312
Db 61 GGACGAAGCAGTCCGCGAGCGCGCGAGCTGGCCCGGGAAGCGGATGCGGTGTTGTC 120
QY 1313 GCGCGGCTGCGGAGCGCTGGACATGTTCAAGAACTTCGAAGAACGCGGACGCTTTCGC 1372
Db 121 GCGCGGCTGCGGAGCGCTGGACATGTTCAAGAACTTCGAAGAACGCGGACGCTTTCGC 180
QY 1373 CAAAGCCGTAGAGGAGCTAGCGTGTGTCGCGTGTGCGCCCTTCCCGTGGCGGTG 1432
Db 181 CAAAGCCGTAGAGGAGCTAGCGTGTGTCGCGTGTGCGCCCTTCCCGTGGCGGTG 240
QY 1433 TTGAGCGCGCACGGCATC 1450
Db 241 TTGAGCGCGCACGGCATC 258

```

RESULT 6  
 US-09-252-991A-7552  
 ; Sequence 7552, Application US/09252991A  
 ; Patent No. 6551795  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Marc J. Rubenfield et al.  
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
 ; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
 ; FILE REFERENCE: 107196.136  
 ; CURRENT APPLICATION NUMBER: US/09/252,991A  
 ; CURRENT FILING DATE: 1999-02-18  
 ; PRIOR APPLICATION NUMBER: US 60/074,788  
 ; PRIOR FILING DATE: 1998-02-18  
 ; PRIOR APPLICATION NUMBER: US 60/094,190  
 ; PRIOR FILING DATE: 1998-07-27  
 ; NUMBER OF SEQ ID NOS: 33142  
 ; SEQ ID NO 7552  
 ; LENGTH: 360  
 ; TYPE: DNA  
 ; ORGANISM: Pseudomonas aeruginosa  
 US-09-252-991A-7552

```

Query Match 11.0%; Score 160; DB 4; Length 360;
Best Local Similarity 99.5%; Pred. No. 2.3e-64;
Matches 210; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 61 TCGCCTCGGACCACTTCCGCATCGTTGCGGCTCGGCAAGAGCGCATGTCCTGGTGC 120
Db 1 TCGCCTCGGACCACTTCCGCATCGTTGCGGCTCGGCAAGAGCGCATGTCCTGGTGC 60
QY 121 GCTACCTGGCGCGCGCGCTTTCCTTTGCGCGTGGTGCATACCCGAGAGAACCCGCCG 180
Db 61 GCTACCTGGCGCGCGCGCTTTCCTTTGCGCGTGGTGCATACCCGAGAGAACCCGCCG 120
QY 181 AGCTGGCACCTCGCTGCGCCAGTATCCGACGTTGGAAGTGGTGGCGGCAACTCGAGC 240
Db 121 AGCTGGCACCTCGCTGCGCCAGTATCCGACGTTGGAAGTGGTGGCGGCAACTCGAGC 180
QY 241 CCGAGTTCTCTGCTCGCGCGCGCGCAACTCTA 271
Db 181 CCGAGTTCTCTGCTCGCGCGCGCGCAACTCTA 211

```

RESULT 7  
 US-09-252-991A-7929/c  
 ; Sequence 7929, Application US/09252991A  
 ; Patent No. 6551795  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Marc J. Rubenfield et al.  
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
 ; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
 ; FILE REFERENCE: 107196.136  
 ; CURRENT APPLICATION NUMBER: US/09/252,991A  
 ; CURRENT FILING DATE: 1999-02-18  
 ; PRIOR APPLICATION NUMBER: US 60/074,788  
 ; PRIOR FILING DATE: 1998-02-18  
 ; PRIOR APPLICATION NUMBER: US 60/094,190  
 ; PRIOR FILING DATE: 1998-07-27  
 ; NUMBER OF SEQ ID NOS: 33142  
 ; SEQ ID NO 7929  
 ; LENGTH: 1170  
 ; TYPE: DNA  
 ; ORGANISM: Pseudomonas aeruginosa  
 US-09-252-991A-7929

```

Query Match 8.3%; Score 120; DB 4; Length 1170;
Best Local Similarity 100.0%; Pred. No. 5.5e-46;
Matches 120; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GTGCTGATCGGCTCGCCACCTTGAAGCTGGTGGAGGACGAGAGCATGAGCCTGAT 61
Db 120 GTGCTGATCGGCTCGCCACCTTGAAGCTGGTGGAGGACGAGAGCATGAGCCTGAT 61

```



```
QY 62 CGCCTCCGACCACTCCGCATCGTTGTGGCTCGGCAAGAGCGGCATGTCCCTGGTGGC 121
|||||
Db 60 CGCCTCCGACCACTCCGCATCGTTGTGGCTCGGCAAGAGCGGCATGTCCCTGGTGGC 1

RESULT 8
US-09-252-991A-7860/c
; Sequence 7860, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 7860
; LENGTH: 1287
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-7860

Query Match 4.3%; Score 62; DB 4; Length 1287;
Best Local Similarity 100.0%; Pred. No. 2.7e-19;
Matches 62; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1389 CTAGCGTGATGCTGTCGCTGTGGCCCTTCCGTCGCCGCTGTGAGCGGCACGGCA 1448
|||||
Db 1287 CTAGCGTGATGCTGTCGCTGTGGCCCTTCCGTCGCCGCTGTGAGCGGCACGGCA 1228

QY 1449 TC 1450
||
Db 1227 TC 1226

SULT 9
-09-252-991A-7623
; Sequence 7623, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 7623
; LENGTH: 1404
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-7623

Query Match 2.5%; Score 36; DB 4; Length 1404;
Best Local Similarity 100.0%; Pred. No. 2.5e-07;
Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GTGCTGATCGGCTCGCCACCTTGAAGCTGCGTTGA 37
|||||
Db 1369 GTGCTGATCGGCTCGCCACCTTGAAGCTGCGTTGA 1404

RESULT 10
US-09-252-991A-9277
```

```
; Sequence 9277, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 9277
; LENGTH: 204
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-9277

Query Match 1.3%; Score 19; DB 4; Length 204;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 821 CCAGTTCGACAAAGCTGCTG 839
|||||
Db 138 CCAGTTCGACAAAGCTGCTG 156

RESULT 11
US-09-252-991A-1743/c
; Sequence 1743, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 1743
; LENGTH: 405
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-1743

Query Match 1.3%; Score 19; DB 4; Length 405;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1040 CGGCGCGCCCTGCGCGC 1058
|||||
Db 98 CGGCGCGCCCTGCGCGC 80

RESULT 12
US-09-252-991A-12452/c
; Sequence 12452, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
```

; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 12452  
; LENGTH: 414  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-12452

Query Match 1.3%; Score 19; DB 4; Length 414;  
Best Local Similarity 100.0%; Pred. No. 18;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 821 CCAGTTCGACAACTGCTG 839  
|||||  
Db 255 CCAGTTCGACAACTGCTG 237

## RESULT 13

US-09-252-991A-8304  
; Sequence 8304, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; PRIOR FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 8304  
; LENGTH: 489  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-8304

Query Match 1.3%; Score 19; DB 4; Length 489;  
Best Local Similarity 100.0%; Pred. No. 18;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 821 CCAGTTCGACAACTGCTG 839  
|||||  
Db 423 CCAGTTCGACAACTGCTG 441

## RESULT 14

US-09-072-596-263  
; Sequence 263, Application US/09072596  
; Patent No. 6458366  
; GENERAL INFORMATION:  
; APPLICANT: Reed, Steven G.  
; APPLICANT: Skeiky, Yasir A.W.  
; APPLICANT: Dillon, Davin C.  
; APPLICANT: Campos-Neto, Antonia  
; APPLICANT: Houghton, Raymond  
; APPLICANT: Vedvick, Thomas S.  
; APPLICANT: Twardzik, Daniel R.  
; APPLICANT: Lodes, Michael J.  
; APPLICANT: Hendrickson, Ronald C.  
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF TUBERCULOSIS  
; NUMBER OF SEQUENCES: 350  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: SEED AND BERRY LLP  
; STREET: 6300 Columbia Center, 701 Fifth Avenue  
; CITY: Seattle  
; STATE: Washington  
; COUNTRY: USA  
; ZIP: 98104-7092  
; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/072,596  
; FILING DATE: 05-MAY-1998  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Maki, David J.  
; REGISTRATION NUMBER: 31,392  
; REFERENCE/DOCKET NUMBER: 210121.417C9  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (206) 622-4900  
; TELEFAX: (206) 682-6031  
; INFORMATION FOR SEQ ID NO: 263:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 522 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
US-09-072-596-263

Query Match 1.3%; Score 19; DB 4; Length 522;  
Best Local Similarity 100.0%; Pred. No. 18;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1082 CGACGGCAAGCTGGTCTG 1100  
|||||  
Db 429 CGACGGCAAGCTGGTCTG 447

## RESULT 15

US-09-252-991A-13195  
; Sequence 13195, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 13195  
; LENGTH: 723  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-13195

Query Match 1.3%; Score 19; DB 4; Length 723;  
Best Local Similarity 100.0%; Pred. No. 17;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 821 CCAGTTCGACAACTGCTG 839  
|||||  
Db 216 CCAGTTCGACAACTGCTG 234

Search completed: August 14, 2003, 07:06:19  
Job time : 117 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: August 14, 2003, 09:25:27 ; Search time 91 Seconds  
(without alignments)  
2172.963 Million cell updates/sec

Title: US-09-701-229-2  
Effect score: 448  
Sequence: 1 MSLASHDFRIIVVLGKSGM.....MFKNFERGLFAKAVEELA 448

Scoring table:  
OLIGO  
Xgapop 60.0 , Xgapext 60.0  
Ygapop 60.0 , Ygapext 60.0  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 569978 seqs, 220691566 residues

Word size: 1

Total number of hits satisfying chosen parameters: 712408

Minimum DB seq length: 25

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlp  
-Q/cgn2\_1/USPTO.spool/US09701229/runat\_09082003.161237.5050/app\_query.fasta.1.647  
-DB=Issued\_Patents\_NA -QFWT-fastap -SUFFIX=olip2n.rn1 -MINMATCH=0.1 -LOOPL-0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi  
-LIST=45 -DOCALLIGN=200 -THR\_SCORE=quality -THR\_MIN=1 -ALIGN=15 -MODE=LOCAL  
-OUTFWT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=25 -MAXLEN=2000000000  
-USER=US09701229.ecgn.1.1.85.grunat\_09082003.161237.5050 -NCPU=3  
-NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6  
-FGAPEXT=7 -FGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database :

- Issued\_Patents\_NA:\*
- 1: /cgn2\_6/ptodata/1/ina/5A\_COMB.seq:\*
- 2: /cgn2\_6/ptodata/1/ina/5B\_COMB.seq:\*
- 3: /cgn2\_6/ptodata/1/ina/6A\_COMB.seq:\*
- 4: /cgn2\_6/ptodata/1/ina/6B\_COMB.seq:\*
- 5: /cgn2\_6/ptodata/1/ina/PTUS\_COMB.seq:\*
- 6: /cgn2\_6/ptodata/1/ina/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	448	100.0	1371	4	US-09-252-991A-7702
C 2	446	99.6	1401	4	Sequence 7702, Ap
C 3	188	42.0	557	4	Sequence 7861, Ap
C 4	127	28.3	564	4	Sequence 7787, Ap
C 5	119	26.6	360	4	Sequence 7928, Ap
C 6	67	15.0	1359	4	Sequence 7522, Ap
C 7	23	5.1	1170	4	Sequence 7624, Ap
C 8	17	3.8	1416	4	Sequence 7929, Ap
C 9	14	3.1	1830121	4	Sequence 1282, Ap
C 10	14	3.1	1830121	4	Sequence 1, Appli
C 11	9	2.0	297	4	Sequence 6271, Ap
C 12	9	2.0	298	4	Sequence 5881, Ap

13	9	2.0	717	4	US-08-936-165A-78	Sequence 78, Appl
14	9	2.0	1242	3	US-09-147-928-1	Sequence 1, Appli
15	9	2.0	1372	4	US-09-530-836-1	Sequence 1, Appli
16	9	2.0	1374	4	US-09-134-001C-1001	Sequence 1001, Ap
17	9	2.0	2423	3	US-08-714-918-86	Sequence 86, Appl
18	9	2.0	2423	3	US-09-265-315-86	Sequence 86, Appl
19	9	2.0	2423	3	US-09-265-315-86	Sequence 86, Appl
20	9	2.0	2423	3	US-09-266-417-86	Sequence 86, Appl
C 21	8	1.8	506	2	US-08-799-173A-15	Sequence 15, Appl
C 22	8	1.8	529	4	US-09-109-204-28	Sequence 28, Appl
C 23	8	1.8	529	4	US-09-490-032-28	Sequence 28, Appl
C 24	8	1.8	655	3	US-09-109-204-27	Sequence 27, Appl
C 25	8	1.8	655	3	US-09-490-032-27	Sequence 27, Appl
C 26	8	1.8	855	4	US-09-252-991A-13763	Sequence 13763, A
C 27	8	1.8	873	4	US-09-252-991A-13662	Sequence 13662, A
C 28	8	1.8	963	4	US-09-252-991A-6979	Sequence 6979, Ap
C 29	8	1.8	990	4	US-09-252-991A-13865	Sequence 13865, A
C 30	8	1.8	1105	2	US-08-799-173A-1	Sequence 1, Appli
C 31	8	1.8	1131	4	US-09-252-991A-7006	Sequence 7006, Ap
C 32	8	1.8	1164	4	US-09-252-991A-1310	Sequence 1310, Ap
C 33	8	1.8	1326	4	US-09-252-991A-1156	Sequence 1156, Ap
C 34	8	1.8	1419	4	US-09-252-991A-1246	Sequence 1246, Ap
C 35	8	1.8	1428	4	US-09-252-991A-13671	Sequence 13671, A
C 36	8	1.8	1557	4	US-09-252-991A-7033	Sequence 7033, Ap
C 37	8	1.8	1779	3	US-09-371-696-1	Sequence 1, Appli
C 38	8	1.8	1886	4	US-09-484-9708-127	Sequence 127, App
C 39	8	1.8	2028	4	US-09-252-991A-7091	Sequence 7091, Ap
C 40	8	1.8	2224	3	US-09-109-204-6	Sequence 6, Appli
C 41	8	1.8	2224	3	US-09-490-032-6	Sequence 6, Appli
C 42	8	1.8	2360	1	US-08-466-390-1	Sequence 1, Appli
C 43	8	1.8	2360	1	US-08-470-950-1	Sequence 1, Appli
C 44	8	1.8	2360	1	US-08-467-781-1	Sequence 1, Appli
C 45	8	1.8	2360	1	US-08-195-487-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1  
US-09-252-991A-7702  
; Sequence 7702, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; PRIOR FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 7702  
; LENGTH: 1371  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-7702

Alignment Scores:  
Pred. No.: 0 Length: 1371  
Score: 448.00 Matches: 448  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 4 Gaps: 0

US-09-701-229-2 (1-448) x US-09-252-991A-7702 (1-1371)

QY 1 MetSerLeuIleAlaSerAspHisPheArgIleValValGlyLeuGlyLysSerGlyMet 20  
|||||  
DB 25 ATGAGCGCTGATCGCTCCGACCACTCCGCGCATGTTCGCGCTCGGCAAGCGCATG 84

QY 21 SerLeuValArgTyrLeuAlaArgArgGlyLeuProPheAlaValValAspThrArgGlu 40  
DB 85 TCCCTGGTGGCTACTGCGCGCGCGCGCTTGCCTTCGCGTGTGCATACCGAGAG 144  
QY 41 AsnProGluLeuAlaThrLeuArgAlaGlnTyrProGlnValGluValArgCysGly 60  
DB 145 AACCCGCGGAGCTGGCCACCTGCTGCGCCAGTATCCGACGAGTGGAGTGGTGGCGC 204  
QY 61 GluLeuAspAlaGluPheLeuCysSerAlaArgGluLeuTyrValSerProGlyLeuSer 80  
DB 205 GAACTCGAGCGCGAGTCTCTGCTCCGCCCGGAACTCTAGCTCAGCGCCCGCTGTGCG 264  
QY 81 LeuArgThrProAlaLeuValGlnAlaAlaLysGlyValArgIleSerGlyAspIle 100  
DB 265 CTGCGCACCCCGCGCTGTGTACAGCGCGCGGAAAGTGTGCGCATCTCCGGTGACATC 324  
QY 101 AspLeuPheAlaArgGluAlaLysAlaProIleValAlaIleThrGlySerAsnAlaLys 120  
DB 325 GATCTCTTCGCGCGGAGCGGAAGCCCGCATGCTGCGCATCACCGGTTCCAAACGGAAG 384  
QY 121 SerThrValThrThrLeuValGlyGluMetAlaValAlaAlaAspLysArgValAlaVal 140  
DB 385 AGCACCGTGACACCTGTGGCGGAATGGCGTGGCGCGGACAAAGCTGTGCCCTC 444  
QY 141 GlyGlyAsnLeuGlyThrProAlaLeuAspLeuLeuAlaAspIleGluLeuTyrVal 160  
DB 445 GCGCGCAACCTGGCACCCCGCGCTCGACCTGTGCGCGGACGACATCGAGCTGTACGTG 504  
QY 161 LeuGluLeuSerSerPheGlnLeuGluThrCysAspArgLeuAsnAlaGluValAlaThr 180  
DB 505 TTGGAGCTGTGAGCTTCAGCTGGAAACCTGCGATCCCTCAACCGCGAGTGGCGACC 564  
QY 181 ValLeuAsnValSerGluAspHisMetAspArgTyrAspGlyMetAlaAspTyrHisLeu 200  
DB 565 GTGCTGAACCTCAGCGAAGACCATATGGATCGTACGACGATGCTGCTACACCTG 624  
QY 201 AlaLysHisArgIlePheArgGlyAlaArgGlnValValAlaAsnArgAlaAspAlaLeu 220  
DB 625 GCCAAGCACCGGATCTTCGCGGGTGGCGCGGAGTGGTGGTGAATCGCGCGCATGCGCGT 684  
QY 221 ThrArgProLeuIleAlaAspThrValProCysTyrPhePheGlyLeuAsnLysProAsp 240  
DB 685 ACCCGACCGCTGATCGCGGATACCGTGGCTGGTGGTGGTGGTGGTGGTGGTGGTGGT 744  
QY 241 PheLysAlaPheGlyLeuIleGluAlaAspGlyGlnLysTyrLeuAlaPheGlnPheAsp 260  
DB 745 TTCAAGGCTTTCGCGCTGATCGAGGACGCGGAGAGTGGCTGGCTGGCTGGCTGGCTGGC 804  
QY 261 LysLeuLeuProValGlyGluLeuLysIleArgGlyAlaHisAsnTyrSerAsnAlaLeu 280  
DB 805 AAGCTGCTGCGCGGTGGCGAATGAAGATCCGCGCGCGGCGGCGGCGGCGGCGGCGGCTC 864  
QY 281 AlaAlaLeuAlaLeuGlyHisAlaValGlyLeuProPheAspAlaMetLeuGlyAlaLeu 300  
DB 865 GCGCGCTGGCGTGGCGCATGCGTGGCGCTGGCGCTGGCGCTGGCGCTGGCGCTGGCGCTG 924  
QY 301 LysAlaPheSerGlyLeuAlaHisArgCysGlnTyrValArgGluArgGlnGlyValSer 320  
DB 925 AAGCGCTTTCGCGCTGGCTGATCGCTGCGAGTGGTGGTGGTGGTGGTGGTGGTGGTGG 984  
QY 321 TyrTyrAspSerLysAlaThrAsnValGlyAlaAlaLeuAlaIleGluGlyLeu 340  
DB 985 TACTACGACGATTCAGCGCCACCAACGTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTG 1044  
QY 341 GlyAlaAspIleAspGlyLysLeuValLeuLeuAlaGlyGlyAspGlyLysGlyAlaAsp 360  
DB 1045 GGTGCGGACATCGAGCGCAAGCTGTGCTGCTGCGCGCGGAGACGCGAGGCGCGCGAT 1104  
QY 361 PheHisAspLeuArgGluProValAlaArgPheCysArgAlaValValLeuLeuGlyArg 380  
DB 1105 TTCCATGACCTGCGCGAGCGCGCTTCGCGCGGCGGTTCGCGCGGCGGTGCTGCTGGCGCGT 1164  
QY 381 AspAlaGlyLeuIleAlaGlnAlaLeuGlyAsnAlaValProLeuValArgValAlaThr 400

DB 1165 GACGCGGCGCTGATTCGCCAGGCACTGGGCAACGCGGTACCGCTGCTGGCTGCGCAACG 1224  
QY 401 LeuAspGluAlaValArgGlnAlaAlaGluLeuAlaArgGluGlyAspAlaValLeuLeu 420  
DB 1225 CTGGACGAAGCAGTCCGCGAGCGCGCGAGCTGGCGCGGAGGCGATGCGGTGCTGTG 1284  
QY 421 SerProAlaCysAlaSerLeuAspMetPheLysAsnPheGluGluArgGlyArgLeuPhe 440  
DB 1285 TCGCGGCGCTGCGCGAGCTGGACATGTTCAAGAACTTCGAAGAACTTCGAAGAACTTC 1344  
QY 441 AlaLysAlaValGluGluLeuAla 448  
DB 1345 GCCAAAGCGTAGAGGAGTAGCG 1368  
RESULT 2  
US-09-252-991A-7861/c  
; Sequence 7861, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252.991A  
; PRIOR FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 7861  
; LENGTH: 1401  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-7861  
Alignment Scores:  
Pred. No.: 0 Length: 1401  
Score: 446.00 Matches: 446  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 99.55% Indels: 0  
DB: 4 Gaps: 0  
US-09-701-229-2 (1-448) x US-09-252-991A-7861 (1-1401)  
QY 1 MetSerLeuIleAlaSerAspHisPheArgIleValValGlyLeuGlyLysSerGlyMet 20  
DB 1338 ATGAGCTGATCGCCTCGACCACTTCGCGATCTGTTCGGCCCTCGGCAAGAGCGGCATG 1279  
QY 21 SerLeuValArgTyrLeuAlaArgArgGlyLeuProPheAlaValValAspThrArgGlu 40  
DB 1278 TCCCTGGTGGCTACTTGGCGCGCGCGGCTTGGCTTTCGCGCTGTGTCGATACCGAGAG 1219  
QY 41 AsnProGluLeuAlaThrLeuArgAlaGlnTyrProGlnValGluValArgCysGly 60  
DB 1218 AACCCGCGGAGCTGGCGACCTCGTGGCGCGGATATCGCGAGTGGAGTGGCTGGCGGC 1159  
QY 61 GluLeuAspAlaGluPheLeuCysSerAlaArgGluLeuTyrValSerProGlyLeuSer 80  
DB 1158 GAACCTGACGCGGAGTTCCTCTGCTCGCGCGGCACTCTACGTCACCGCGCGCTGTGCG 1099  
QY 81 LeuArgThrProAlaLeuValGlnAlaAlaLysGlyValArgIleSerGlyAspIle 100  
DB 1098 CTGCGCACCCCGCGCTGTGTACAGCGCGCGGAGAGTGTGCGCATCTCCGGTGACATC 1039  
QY 101 AspLeuPheAlaArgGluAlaLysAlaProIleValAlaIleThrGlySerAsnAlaLys 120  
DB 1038 GATCTCTTCGCGCGGAGCGGCGGCGCGATCTGCTGCCCATCACCGGTTCCACGCGAAG 979  
QY 121 SerThrValThrThrLeuValGlyGluMetAlaValAlaAlaAspLysArgValAlaVal 140



```

; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 7928
; LENGTH: 564
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-7928

```

Alignment Scores:	1.26e-115	Length:	564
Pred. No.:	127.00	Matches:	127
Score:	100.00%	Conservative:	0
Percent Similarity:	100.00%	Mismatches:	0
Best Local Similarity:	100.00%	Indels:	0
Query Match:	28.35%	Gaps:	0
DB:	4		

US-09-701-229-2 (1-448) x US-09-252-991A-7928 (1-564)

322	TyrAspAspSerLysAlaThrAsnValGlyAlaAlaLeuAlaIleGluGlyLeuGly	341
563	TACGACGATTCCAAAGGCCACCAACGTCGGCGCCCTGCGCGCATCGAGGGCGTGGCT	504
342	AlaAspIleAspGlyLysLeuValLeuAlaGlyClyAspGlyLysGlyAlaAspPhe	361
503	GCGGACATCGACGCCAAGCTGGTCTGTCGCGCGGACGCGAAGCGCGCGATTTC	444
362	HisAspLeuArgGluProValAlaArgPheCysArgAlaValValLeuLeuGlyArgAsp	381
443	CATGACCTGCGCGAGCGCGTCCCGCGTTCGCGGCGGTGTACTGCTTGGCCGTGAC	384
382	AlaGlyLeuIleAlaGlnAlaLeuGlyAsnAlaValProLeuValArgValAlaThrLeu	401
383	GCGGGCTGATTGCCAGGCACTGGCAACGCGTACCGCTGCTGCGCTGCACAGCTG	324
402	AspGluAlaValArgGlnAlaAlaGluLeuAlaArgGluClyAspAlaValLeuLeuSer	421
323	GACGAAGCAGTCCCGCAGGCGCGACGTGCGCGCAAGCGCATCGGTGTGTGTGCG	264
422	ProAlaCysAlaSerLeuAspMetPheLysAsnPhelGluArgGlyArgLeuPheAla	441
263	CGGCGCTCGCGAGCCTGGACATGTTCCAGNACTTCGAAGACCGCGCCTGTTCGCC	204
442	LysAlaValGluLeuAla	448
203	AAAGCCGTAGAGGAGCTAGCG	183

RESULT 5  
US-09-252-991A-7552  
; Sequence 752, Application US/09252991A  
. Patent No. 6551795  
GENERAL INFORMATION:

APPLICANTI: MARC J. KUDENHLEID ET AL.  
 TITOLO DI INVENZIONE: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
 ;  
 TITOLO DI INVENZIONE: AERGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
 ;

Alignment Scores:  
Pred. No.:  
Score:  
Percent Similarity:  
Best Local Similarity:  
Query Match:  
DB:

US-09-701-229-2 (1-448) x US-09-252-991A-7552 (1-360)

QY	5	AlaSerAspHisPheArgIleValValGlyLeuGlyLySerGlyMetSerLeuValArg	24
DB	3	GCCTCCGACCACTCCGCATCGTTGCGGCTCGGCAAGACGGCATGTCCCTGGTGGCC	62
QY	25	TyrLeuAlaArgArgGlyLeuProPheAlaValValAspThrArgGluAsnProProGlu	44
DB	63	TACCTGGCGCGCGCGGCTTGCCCTTCGCGCTTGTGCATACCCGAGAGAACCGCGCGAG	122
QY	45	LeuAlaThrLeuArgAlaGlnTyrProGlnValGluValArgCysGlyGluLeuAspAla	64
DB	123	CTGGGCCACCTCGTGGGCCAGTATCCGAGGTGAGTGGTGTGCGCGCAACTGCACGCC	182
QY	65	GluPheLeuCysSerAlaArgGluLeuTyrValSerProGlyLeuSerLeuArgThrPro	84
DB	183	GAGTTCCTCTGTCTCGCGCGCGAACTACGTCAAGCCCGCATGTGCTGTCGCACCCGC	242
QY	85	AlaLeuValGlnAlaAlaLysGlyValArgIleSerGlyAspIleAspLeuPheAla	104
DB	243	GGCTGGTACAGCGCGCGGAAGGTGTGGCATCTCCGGTGACATCGATCTCTTCGCC	302
QY	105	ArgGluAlaLysAlaProIleValAlaIleThrGlySerAsnAlaLysSerThrVal	123
DB	303	CCGAGGCGAAGGCCCATCGTCGCATCACCGTTCCAAACGCGAAGACACCGTG	359

## RESULT 6

US-09-252-991A-7624  
; Sequence 7624, Application US/09252991A  
; Patent No. 6551795

; GENERAL INFORMATION: Rubenfield et al.  
 ; APPLICANT: Marc J. NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
 ; TITLE OF INVENTION: ARBUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
 ; TITLE OF INVENTION: ARBUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

Alignment Scores:	2.22e-56	Length:	1359
Pred. No.:	67.00	Matches:	67
Score:	100.00%	Conservative:	0
Percent Similarity:	100.00%	Mismatches:	0
Best Local Similarity:	100.00%	Indels:	0
Query Match:	14.96%	Gaps:	0
DB:	4		

US-09-701-229-2 (1-448) x US-09-252-991A-7624 (1-1359)

QY	382	AlaGlyLeuIleAlaGlnAlaLeuGlyAsnAlaValProLeuValArgValAlaThrLeu	401
Db	2	GCCGGGCTGATTGCCCAAGGCACCTGGGCAAGCGGTACCGCTGTGGCGCTGCACAGCTG	61
QY	402	AspGluAlaValArgGlnAlaAlaGluLeuAlaArgGluGlyAspAlaValLeuLeuSer	421

Db 62 GACGAAGACAGTCCGGCAGCGCCGAGCTGGCCCGGCGAAGCGGCGTGTGTCTG 121  
 QY 422 ProAlaCysAlaSerLeuAspMetPheLysAsnPhelGluArgGlyAlaGLeuPheAla 441  
 Db 122 CCGGCTGCGCGAGCTGGACATGTTCAAGAACTTCGAAGACCGGACGCCCTGTTCGCC 181  
 QY 442 LysAlaValGluGluLeuAla 448  
 Db 182 AAAGCCGTAGAGGAGCTAGCG 202

RESULT 7  
 US-09-252-991A-7929/c  
 ; Sequence 729, Application US/09252991A  
 ; Patent No. 6551795  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Marc J. Rubenfield et al.  
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
 ; FILE REFERENCE: 107196.136  
 ; CURRENT APPLICATION NUMBER: US/09/252,991A  
 ; PRIOR FILING DATE: 1999-02-18  
 ; PRIOR FILING DATE: 1998-02-18  
 ; PRIOR FILING DATE: 1998-02-18  
 ; PRIOR FILING DATE: 1998-02-18  
 ; PRIOR FILING DATE: 1998-07-27  
 ; NUMBER OF SEQ ID NOS: 33142  
 ; SEQ ID NO 7929  
 ; LENGTH: 1170  
 ; ORGANISM: Pseudomonas aeruginosa  
 US-09-252-991A-7929

Alignment Scores:  
 Pred No.: 2,92e-13 Length: 1170  
 Score: 23.00 Matches: 23  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 5.13% Indels: 0  
 DB: 4 Gaps: 0

US-09-701-229-2 (1-448) x US-09-252-991A-7929 (1-1170)  
 QY 1 MetSerLeuIleAlaSerAspHisPheArgIleValValGlyLeuGlyLysSerGlyMet 20  
 Db 71 ATGACGCTGATCGCTCCGACCATCTCCGATCGTTCGGCGCGGCAAGCGGCATG 12  
 QY 21 SerLeuVal 23  
 Db 11 TCCCTGGTG 3

RESULT 8  
 US-09-328-352-1282  
 ; Sequence 1282, Application US/09328352  
 ; Patent No. 6562958  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Gary L. Breton et al.  
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER  
 ; FILE REFERENCE: GTC99-03PA  
 ; CURRENT APPLICATION NUMBER: US/09/328,352  
 ; CURRENT FILING DATE: 1999-06-04  
 ; NUMBER OF SEQ ID NOS: 8252  
 ; SEQ ID NO 1282  
 ; LENGTH: 1416  
 ; TYPE: DNA  
 ; ORGANISM: Acinetobacter baumannii  
 US-09-328-352-1282  
 Alignment Scores:  
 Pred No.: 2,7e-07 Length: 1416  
 Score: 17.00 Matches: 17  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0

Query Match: 3.79% Indels: 0  
 DB: 4 Gaps: 0  
 US-09-701-229-2 (1-448) x US-09-328-352-1282 (1-1416)  
 QY 110 ProIleValAlaIleThrGlySerAsnAlaLysSerThrValThrLeu 126  
 Db 385 CCGATTGTGGGATTACAGTTCCAAATGCCAAAGTACTGTAACCACTTTA 435  
 RESULT 9  
 US-09-557-884-1  
 ; Sequence 1, Application US/09557884  
 ; Patent No. 6506581  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Fleischmann et al.  
 ; TITLE OF INVENTION: The nucleotide sequence of the Haemophilus influenzae Rd Genome, Fragments Thereof, and Uses Thereof  
 ; NUMBER OF SEQUENCES: 1  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Human Genome Sciences, Inc.  
 ; STREET: 9410 Key West Avenue  
 ; CITY: Rockville  
 ; STATE: MD  
 ; COUNTRY: USA  
 ; ZIP: 20850  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: 3 1/2 inch diskette  
 ; COMPUTER: Dell Pentium  
 ; OPERATING SYSTEM: MS DOS v6.22  
 ; SOFTWARE: ASCII Text  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/09/557,884  
 ; FILING DATE: 25-Apr-2000  
 ; CLASSIFICATION: <Unknown>  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: 08/476,102  
 ; FILING DATE: JUN-5-1995  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Michelle S. Marks  
 ; REGISTRATION NUMBER: 41,971  
 ; REFERENCE/DOCKET NUMBER: PB186P3  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 301-309-8504  
 ; TELEFAX: 301-309-8439  
 ; INFORMATION FOR SEQ ID NO: 1:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 1830121 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: double  
 ; TOPOLOGY: linear  
 ; SEQUENCE DESCRIPTION: SEQ ID NO: 1:  
 US-09-557-884-1  
 Alignment Scores:  
 Pred No.: 0.212 Length: 1830121  
 Score: 14.00 Matches: 14  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 3.12% Indels: 0  
 DB: 4 Gaps: 0  
 US-09-701-229-2 (1-448) x US-09-557-884-1 (1-1830121)  
 QY 157 GluLeuTyrrValLeuGluLeuSerSerPheGlnLeuGluThr 170  
 Db 1203588 GAACCTTATGACTAGAGCTTCTAGTTTTCAGCTTGAGACA 1203629  
 RESULT 10  
 US-09-643-990A-1  
 ; Sequence 1, Application US/09643990A  
 ; Patent No. 6528289  
 ; GENERAL INFORMATION:



```
; APPLICANT: Robert D. Fleischmann
; Mark D. Adams
; Owen White
; Hamilton O. Smith
; J. Craig Venter
;
; TITLE OF INVENTION: The Nucleotide sequence of
; the Haemophilus influenzae Rd Genome, Fragments
; Thereof, and Uses Thereof
;
; NUMBER OF SEQUENCES: 1
;
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville,
; STATE: MD
; COUNTRY: USA
; ZIP: 20850
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3 1/2 inch diskette
; COMPUTER: Dell Pentium
; OPERATING SYSTEM: MS DOS v6.22
; SOFTWARE: ASCII Text
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/643,990A
; FILING DATE: 23-Aug-2000
; CLASSIFICATION: <Unknown>
;
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/487,429
; FILING DATE: 1995-06-07
; APPLICATION NUMBER: 08/426,787
; FILING DATE: 1995-04-21
; ATTORNEY/AGENT INFORMATION:
; NAME: Kenley K. Hoover
; REGISTRATION NUMBER: 40,302
; REFERENCE/DOCKET NUMBER: PB186PIC1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 301-610-5790
; TELEFAX: 310-309-8439
;
; INFORMATION FOR SEQ ID NO: 1:
;
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1830121 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
;
; SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-643-990A-1

Alignment Scores:
Pred. No.: 0.212 Length: 1830121
Score: 14.00 Matches: 14
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3.12% Indels: 0
DB: 4 Gaps: 0

US-09-701-229-2 (1-448) x US-09-643-990A-1 (1-1830121)
RESULT 11
US-09-313-294A-6271/c
; Sequence 6271, Application US/09313294A
; Patent No. 6476212
; GENERAL INFORMATION:
; APPLICANT: Lalgudi, Raghunath V.
; APPLICANT: Ito, Laura Y.
; APPLICANT: Sherman, Bradley K.
; TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM CORN EAR
; FILE REFERENCE: PL-0017 US
; CURRENT APPLICATION NUMBER: US/09/313,294A
; CURRENT FILING DATE: 1999-05-14
; NUMBER OF SEQ ID NOS: 7600

157 GluLeuTyValLeuGluLeuSerPheGlnLeuGluThr.170
DB 1203588 GAACCTTATGTACTAGAGCTTCTTAGTTTACCTTGAGACA 1203629

US-09-701-229-2 (1-448) x US-09-313-294A-6271 (1-297)
; SOFTWARE: PERL Program
; SEQ ID NO 6271
; LENGTH: 297
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No. 6476212 700351520H1
; NAME/KEY: unsure
; LOCATION: 133-178
; OTHER INFORMATION: a, t, c, g, or other
US-09-313-294A-6271

Alignment Scores:
Pred. No.: 4.37 Length: 297
Score: 9.00 Matches: 9
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 2.01% Indels: 0
DB: 4 Gaps: 0

US-09-701-229-2 (1-448) x US-09-313-294A-6271 (1-297)
QY 348 LeuValLeuAlaGlyAspGly 356
DB 261 CTCGTCCTCTCGCGGTGGTGACGGC 235

RESULT 12
US-09-313-294A-5881/c
; Sequence 5881, Application US/09313294A
; Patent No. 6476212
; GENERAL INFORMATION:
; APPLICANT: Lalgudi, Raghunath V.
; APPLICANT: Ito, Laura Y.
; APPLICANT: Sherman, Bradley K.
; TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM CORN EAR
; FILE REFERENCE: PL-0017 US
; CURRENT APPLICATION NUMBER: US/09/313,294A
; CURRENT FILING DATE: 1999-05-14
; NUMBER OF SEQ ID NOS: 7600
; SOFTWARE: PERL Program
; SEQ ID NO 5881
; LENGTH: 298
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No. 6476212 700350840H1
; NAME/KEY: unsure
; LOCATION: 58, 109-154, 263
; OTHER INFORMATION: a, t, c, g, or other
US-09-313-294A-5881

Alignment Scores:
Pred. No.: 4.38 Length: 298
Score: 9.00 Matches: 9
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 2.01% Indels: 0
DB: 4 Gaps: 0

US-09-701-229-2 (1-448) x US-09-313-294A-5881 (1-298)
QY 348 LeuValLeuAlaGlyAspGly 356
DB 237 CTCGTCCTCTCGCGGTGGTGACGGC 211

RESULT 13
US-08-936-165A-78
; Sequence 78, Application US/08936165A
; Patent No. 6348582
; GENERAL INFORMATION:
; APPLICANT: Black, Michael
```

APPLICANT: Burnham, Martin  
APPLICANT: Hodgson, John  
APPLICANT: Knowles, David  
APPLICANT: Lonetto, Michael  
APPLICANT: Nicholas, Richard  
APPLICANT: Pratt, Julie  
APPLICANT: Reichard, Richard  
APPLICANT: Rosenberg, Martin  
APPLICANT: Ward, Judith  
TITLE OF INVENTION: No. 6348582el Prokaryotic Polynucleotides,  
NUMBER OF INVENTION: Polypeptides and Their Uses  
NUMBER OF SEQUENCES: 534  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: SmithKline Beecham Corporation  
STREET: 709 Swedeland Road  
CITY: King of Prussia  
STATE: PA  
COUNTRY: USA  
ZIP: 19406-0939  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/936,165A  
FILING DATE: 24-SEP-1997  
CLASSIFICATION: 536  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 60/027,032  
FILING DATE: 24-SEP-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Gimmi, Edward R  
REGISTRATION NUMBER: 38,891  
REFERENCE/DOCKET NUMBER: P50549  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 610-270-4478  
TELEFAX: 610-270-5090  
TELEX:  
INFORMATION FOR SEQ ID NO: 78:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 717 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: Genomic DNA  
US-08-936-165A-78  
Alignment Scores:  
Pred. No.: 10.1 Length: 717  
Score: 9.00 Matches: 9  
Percent Similarity: 100.00%  
Best Local Similarity: 100.00%  
Query Match: 2.01%  
Indels: 0  
Gaps: 0  
US-09-701-229-2 (1-448) x US-08-936-165A-78 (1-717)  
QY 418 ValLeuLeuSerProAlaCysAlaSer 426  
Db 411 GTATTATTGTCACCTGCTGTGGGAGT 437  
RESULT 14  
US-09-147-928-1  
Sequence 1, Application US/09147928  
Patent No. 6211161  
GENERAL INFORMATION:  
APPLICANT: Palmer, Leslie M.  
APPLICANT: Pratt, Julie M.  
APPLICANT: Hodgson, John E.  
APPLICANT: Beattie, David T.  
APPLICANT: Lowe, Adrian M.  
APPLICANT: Lonetto, Michael A.

APPLICANT: Nicholas, Richard O.  
APPLICANT: Deresiewicz, Robert L.  
TITLE OF INVENTION: MURD  
FILE REFERENCE: GM10106  
CURRENT APPLICATION NUMBER: US/09/147,928  
CURRENT FILING DATE: 1999-03-19  
EARLIER APPLICATION NUMBER: 60/061,064  
EARLIER FILING DATE: 1997-10-03  
NUMBER OF SEQ ID NOS: 6  
SOFTWARE: FastSEQ for Windows Version 3.0  
SEQ ID NO 1  
LENGTH: 1242  
TYPE: DNA  
ORGANISM: Staphylococcus aureus  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (1)...(1239)  
NAME/KEY: misc\_feature  
LOCATION: (1)...(1242)  
OTHER INFORMATION: n = A,T,C or G  
US-09-147-928-1  
Alignment Scores:  
Pred. No.: 17 Length: 1242  
Score: 9.00 Matches: 9  
Percent Similarity: 100.00%  
Best Local Similarity: 100.00%  
Query Match: 2.01%  
Indels: 0  
Gaps: 0  
US-09-701-229-2 (1-448) x US-09-147-928-1 (1-1242)  
QY 418 ValLeuLeuSerProAlaCysAlaSer 426  
Db 1138 GTATTATTGTCACCTGCTGTGGGAGT 1164  
RESULT 15  
US-09-530-836-1  
Sequence 1, Application US/09530836  
Patent No. 6534284  
GENERAL INFORMATION:  
APPLICANT: Mohammed El-Sherbeini  
APPLICANT: Kenny Kin Wong  
APPLICANT: Wayne M. Geissler  
TITLE OF INVENTION: MURD PROTEIN AND GENE OF STAPHYLOCOCCUS  
FILE REFERENCE: 19985P  
CURRENT APPLICATION NUMBER: US/09/530,836  
CURRENT FILING DATE: 2000-05-03  
PRIOR APPLICATION NUMBER: 60/064,183  
PRIOR FILING DATE: 1997-11-04  
PRIOR APPLICATION NUMBER: PCT/US98/23156  
PRIOR FILING DATE: 1998-10-30  
NUMBER OF SEQ ID NOS: 6  
SOFTWARE: FastSEQ for Windows Version 4.0  
SEQ ID NO 1  
LENGTH: 1372  
TYPE: DNA  
ORGANISM: Staphylococcus aureus murd  
US-09-530-836-1  
Alignment Scores:  
Pred. No.: 18.6 Length: 1372  
Score: 9.00 Matches: 9  
Percent Similarity: 100.00%  
Best Local Similarity: 100.00%  
Query Match: 2.01%  
Indels: 0  
Gaps: 0  
US-09-701-229-2 (1-448) x US-09-530-836-1 (1-1372)  
QY 418 ValLeuLeuSerProAlaCysAlaSer 426  
Db 1138 GTATTATTGTCACCTGCTGTGGGAGT 1164

Db 1258 GTATTATTGTCACCTGCTTGCGAGT 1284

Search completed: August 14, 2003, 11:33:45  
Job time : 1195 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: August 14, 2003, 09:38:07 ; Search time 350 Seconds

(without alignments)

2864.564 Million cell updates/sec

Title: US-09-701-229-2

Perfect score: 448

Sequence: 1 MSIIASHDFRIVVGLGKSGM.....MFKNFEEGRGLFAKAVEELA 448

Scoring table:

OLIGO  
Xgapop 60.0 , Xgapext 60.0  
Ygapop 60.0 , Ygapext 60.0  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 1504479 seqs, 1118970152 residues

Word size: 1

Total number of hits satisfying chosen parameters: 2626184

Minimum DB seq length: 25

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlp  
-Q=/cgn2\_1/USPO.spool/US09701229/runat\_09082003\_161238\_5080/app.query.fasta\_1.647  
-DB=PublishedApplications\_NA -QFMT=fastap -SUFFIX=olip2n.rnpb -MINMATCH=0.1  
-LOOPCL=0 -LOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=oligo  
-TRANS=human40.cdi -LIST=45 -DOALIGN=200 -THR\_SCORE=quality -THR\_MIN=1  
-ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=25  
-MAXLEN=2000000000 -USER=US09701229\_@CGN\_1\_1\_333\_@runat\_09082003\_161238\_5080  
-NCPU=6 -ICPU=3 -NO\_MMAP -LARGQUERY -NEG\_SCORES=0 -WAIT -DSPBLOCK=100  
-LONGLOG -DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60  
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database :

PublishedApplications\_NA:\*  
1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq:\*  
2: /cgn2\_6/ptodata/1/pubpna/PCT\_NEW\_PUB.seq:\*  
3: /cgn2\_6/ptodata/1/pubpna/US06\_NEW\_PUB.seq:\*  
4: /cgn2\_6/ptodata/1/pubpna/US06\_PUBCOMB.seq:\*  
5: /cgn2\_6/ptodata/1/pubpna/US07\_NEW\_PUB.seq:\*  
6: /cgn2\_6/ptodata/1/pubpna/PCTUS\_PUBCOMB.seq:\*  
7: /cgn2\_6/ptodata/1/pubpna/US08\_NEW\_PUB.seq:\*  
8: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq:\*  
9: /cgn2\_6/ptodata/1/pubpna/US09A\_PUBCOMB.seq:\*  
10: /cgn2\_6/ptodata/1/pubpna/US09B\_PUBCOMB.seq:\*  
11: /cgn2\_6/ptodata/1/pubpna/US09C\_PUBCOMB.seq:\*  
12: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq:\*  
13: /cgn2\_6/ptodata/1/pubpna/US10A\_PUBCOMB.seq:\*  
14: /cgn2\_6/ptodata/1/pubpna/US10B\_PUBCOMB.seq:\*  
15: /cgn2\_6/ptodata/1/pubpna/US10\_NEW\_PUB.seq:\*  
16: /cgn2\_6/ptodata/1/pubpna/US60\_NEW\_PUB.seq:\*  
17: /cgn2\_6/ptodata/1/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Length	ID	Description
1	448	100.0	1347 9	US-09-815-242-7912 Sequence 7912, Ap

2	14	3.1	1314	9	US-09-815-242-7090	Sequence 7090, Ap
3	14	3.1	1317	9	US-09-741-669-259	Sequence 259, App
4	14	3.1	1317	9	US-09-815-242-5932	Sequence 5932, Ap
5	14	3.1	1317	9	US-09-815-242-9978	Sequence 9978, Ap
6	14	3.1	1830121	14	US-10-323-960-1	Sequence 1, Appli
c 7	9	2.0	135	9	US-09-815-242-3503	Sequence 3503, Ap
c 8	9	2.0	135	9	US-09-815-242-3597	Sequence 3597, Ap
c 9	9	2.0	216	9	US-09-815-242-1519	Sequence 1519, Ap
c 10	9	2.0	327	9	US-09-815-242-1487	Sequence 1487, Ap
c 11	9	2.0	383	9	US-09-815-242-2986	Sequence 2986, Ap
c 12	9	2.0	429	13	US-10-027-632-180203	Sequence 180203,
c 13	9	2.0	429	13	US-10-027-632-180204	Sequence 180204,
c 14	9	2.0	485	9	US-09-815-242-2037	Sequence 2037, Ap
c 15	9	2.0	498	13	US-10-027-632-280432	Sequence 280432,
c 16	9	2.0	710	13	US-10-027-632-14190	Sequence 14190, A
c 17	9	2.0	717	9	US-09-939-980-78	Sequence 78, Appl
c 18	9	2.0	719	13	US-10-027-632-16087	Sequence 16087, A
c 19	9	2.0	999	12	US-10-259-165-369	Sequence 369, App
c 20	9	2.0	1002	12	US-10-259-165-17	Sequence 17, Appl
c 21	9	2.0	1320	8	US-08-781-986A-80	Sequence 80, Appl
c 22	9	2.0	1347	9	US-09-815-242-4419	Sequence 4419, Ap
c 23	9	2.0	1350	9	US-09-815-242-8250	Sequence 8250, Ap
c 24	9	2.0	1350	9	US-09-815-242-8695	Sequence 8695, Ap
c 25	9	2.0	1491	14	US-10-156-761-3569	Sequence 3569, Ap
c 26	9	2.0	9025608	14	US-10-156-761-1	Sequence 1, Appli
c 27	8	1.8	25	14	US-10-098-263B-120644	Sequence 120644,
c 28	8	1.8	271	9	US-09-815-242-878	Sequence 878, App
c 29	8	1.8	282	14	US-10-156-761-5138	Sequence 5138, Ap
c 30	8	1.8	338	9	US-09-864-761-31989	Sequence 31989, A
c 31	8	1.8	370	13	US-10-027-632-68104	Sequence 68104, A
c 32	8	1.8	370	13	US-10-027-632-68105	Sequence 68105, A
c 33	8	1.8	388	10	US-09-960-352-41	Sequence 41, Appl
c 34	8	1.8	403	10	US-09-960-352-5964	Sequence 5964, Ap
c 35	8	1.8	403	10	US-09-960-352-11363	Sequence 11363, A
c 36	8	1.8	404	13	US-10-027-632-39955	Sequence 39955, A
c 37	8	1.8	404	13	US-10-027-632-39956	Sequence 39956, A
c 38	8	1.8	419	10	US-09-960-352-969	Sequence 969, App
c 39	8	1.8	467	11	US-09-918-995-21561	Sequence 21561, A
c 40	8	1.8	490	9	US-09-864-761-15475	Sequence 15475, A
c 41	8	1.8	519	13	US-10-027-632-85411	Sequence 85411, A
c 42	8	1.8	519	13	US-10-027-632-314877	Sequence 314877,
c 43	8	1.8	520	13	US-10-027-632-288812	Sequence 288812,
c 44	8	1.8	529	14	US-10-284-985-28	Sequence 28, Appl
c 45	8	1.8	541	9	US-09-864-761-14821	Sequence 14821, A

#### ALIGNMENTS

#### RESULT 1

US-09-815-242-7912  
Sequence 7912, Application US/09815242  
Patent No. US20020061569A1

#### GENERAL INFORMATION:

APPLICANT: Haselbeck, Robert  
APPLICANT: Ohlsen, Kari L.  
APPLICANT: Zyskind, Judith W.  
APPLICANT: Wall, Daniel  
APPLICANT: Trawick, John D.  
APPLICANT: Carr, Grant J.  
APPLICANT: Yamamoto, Robert T.  
APPLICANT: Xu, H. Howard  
TITLE OF INVENTION: Identification of Essential Genes in  
FILE REFERENCE: ELITRA.011A  
CURRENT APPLICATION NUMBER: US/09/815,242  
CURRENT FILING DATE: 2001-03-21  
PRIOR APPLICATION NUMBER: 60/191,078  
PRIOR FILING DATE: 2000-03-21  
PRIOR APPLICATION NUMBER: 60/206,848  
PRIOR FILING DATE: 2000-05-23  
PRIOR APPLICATION NUMBER: 60/207,727  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: 60/242,578

```

: PRIOR FILING DATE: 2000-10-23
: PRIOR APPLICATION NUMBER: 60/253,625
: PRIOR FILING DATE: 2000-11-27
: PRIOR APPLICATION NUMBER: 60/257,931
: PRIOR FILING DATE: 2000-12-22
: PRIOR APPLICATION NUMBER: 60/269,308
: PRIOR FILING DATE: 2001-02-16
: NUMBER OF SEQ ID NOS: 14110
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 7912
: LENGTH: 1347

```

Alignment Scores:			
Pred. No.:	0	Length:	1347
Score:	448.00	Matches:	448
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	9	Gaps:	0

US-09-701-229-2 (1-448) x US-09-815-242-7912 (1-1347)

Qy	1	MetSerLeuIleAlaSerAspHisPheArgIleValValGlyLeuGlyLysSerGlyMet	20
Db	1	ATGAGCTGATCGCTCGGACACTTCGCGATCTTTCGGCTTCGGCAAGACGCGCATG	60
Qy	21	SerLeuValArgTyrLeuAlaArgGlyLeuProPheAlaValValAspThrArgGlu	40
Db	61	TCCCTGGTGGCTACTCGCGCGCGGGCTTGCCTTTCGCCGTGTCGATACCCGAGAG	120
Qy	41	AsnProProGluLeuAlaThrLeuArgAlaGlnTyrProGlnValGluValArgCysGly	60
Db	121	AAACCGCGGAGTGGCCACCCTCGGTGCCAGTATCCGAGGTGGAAGTGCCTTCCGCG	180
Qy	61	GluLeuAspAlaGluPheLeuCysSerAlaArgGluLeuTyrValSerProGlyLeuSer	80
Db	181	GAACTCGACCGCGAGTCTCTGTCTCGCCCGCGAACTCTATGTACGCCCGCTTGTGC	240
Qy	81	LeuArgThrProAlaLeuValGlnAlaAlaLysGlyValArgIleSerGlyAspIle	100
Db	241	CTGCGCACCCCTGCGCTGGTACAGCGCGCCGAAAGCGGTGCGCATCTCCGCTGACATC	300
Qy	101	AspLeuPheAlaArgGluAlaLysAlaProIleValAlaIleThrGlySerAsnAlaLys	120
Db	301	GATCTCTCGCCCGGAGCGAAGCCCGCATCGTCCCATCACCGGTTCACACGGCAAG	360
Qy	121	SerThrValThrLeuValGlyGluMetAlaValAlaAlaAspLysArgValAlaVal	140
Db	361	AGCACCGTGACCACCCCTGTTGGCGAAATGCGGTGCGCGGACAAAGCGTGTGCGCGTC	420
Qy	141	GlyGlyAsnLeuGlyThrProAlaLeuAspLeuLeuAlaAspAspIleGluLeuTyrVal	160
Db	421	GGCGGCACCTCGGCACCCCGCGCTCGACCTGCTGCCGACACATCGAGTGTACGTT	480
Qy	161	LeuGluLeuSerSerPheGlnLeuGluThrCysAspArgLeuAsnAlaGluValAlaThr	180
Db	481	TTGGAGCTGTCGAGCTTCCAGCTGGAAACCTGCGATCGCCTCAACGCCGAGTGGCGACC	540
Qy	181	ValLeuAsnValSerGluAspHisMetAspArgTyrAspGlyMetAlaAspTyrHisLeu	200
Db	541	GTGCTGAACGTTCAGCGAAGACCATATGGATCGGTACGACGCGCATGCTGACTACCACTG	600
Qy	201	AlaLysHisArgIlePheArgGlyAlaArgGlnValValValAsnArgAlaAspAlaLeu	220
Db	601	GCCAGCACCGGAGTCTCCGCGGTGCCCGCAGGTCGTGTGTAATCGCCCGATGCCCTG	660
Qy	221	ThrArgProLeuIleAlaAspThrValProCysTyrSerPheGlyLeuAsnLysProAsp	240

## RESULT 2

```

US-09-815-242-7090
; Sequence 7090, Application US/09815242
; Patent No. US20020081569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyaskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Es
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23

```

```
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 1410
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7090
; LENGTH: 1314
; TYPE: DNA
; ORGANISM: Haemophilus influenzae
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1314)
US-09-815-242-7090

Alignment Scores:
Pred. No.: 0.000202 Length: 1314
Score: 14.00 Matches: 14
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3.12% Indels: 0
DB: 9 Gaps: 0

US-09-701-229-2 (1-448) x US-09-815-242-7090 (1-1314)
QY 157 GluLeuTyrrValLeuGluLeuSerPheGlnLeuGluThr 170
Db 457 GAACCTTATGCTAGAGCTTCTAGTTTTCAGCTTGAGACA 498

RESULT 3
US-09-741-669-259
; Sequence 259, Application US/09741669
; Patent No. US20020022718A1
; GENERAL INFORMATION:
; APPLICANT: Forsyth, R. Allyn
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; TITLE OF INVENTION: Genes identified as required for
; FILE REFERENCE: ELITRA.009A
; CURRENT APPLICATION NUMBER: US/09/741,669
; CURRENT FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: US 60/173005
; PRIOR FILING DATE: 1999-12-23
; NUMBER OF SEQ ID NOS: 481
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 259
; LENGTH: 1317
; TYPE: DNA
; ORGANISM: Escherichia coli
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1317)
US-09-741-669-259

Alignment Scores:
Pred. No.: 0.000203 Length: 1317
Score: 14.00 Matches: 14
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3.12% Indels: 0
DB: 9 Gaps: 0

US-09-701-229-2 (1-448) x US-09-741-669-259 (1-1317)
QY 157 GluLeuTyrrValLeuGluLeuSerPheGlnLeuGluThr 170
Db 457 GAACCTAGCTGCTGGAAGCTGCTGAGCTCCAGCTGGAACC 498

RESULT 4
US-09-815-242-5932
```

```
; Sequence 5932, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 1410
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5932
; LENGTH: 1317
; TYPE: DNA
; ORGANISM: Escherichia coli
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1317)
US-09-815-242-5932

Alignment Scores:
Pred. No.: 0.000203 Length: 1317
Score: 14.00 Matches: 14
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3.12% Indels: 0
DB: 9 Gaps: 0

US-09-701-229-2 (1-448) x US-09-815-242-5932 (1-1317)
QY 157 GluLeuTyrrValLeuGluLeuSerPheGlnLeuGluThr 170
Db 457 GAACCTAGCTGCTGGAAGCTGCTGAGCTCCAGCTGGAACC 498

RESULT 5
US-09-815-242-9978
; Sequence 9978, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
```

```

; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9978
; LENGTH: 1317
; TYPE: DNA
; ORGANISM: Salmonella typhi
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1317)
US-09-815-242-9978

Alignment Scores:
Pred. No.: 0.000203 Length: 1317
Score: 14.00 Matches: 14
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3.12% Indels: 0
DB: 9 Gaps: 0

US-09-701-229-2 (1-448) x US-09-815-242-9978 (1-1317)
QY 157 GluLeuTyValLeuGluLeuSerPheGlnLeuGluThr 170
Db 457 GAATGTACTGCTGGAATATCCAGTTTCCTCCAAATGGAAC 498

RESULT 6
US-10-329-960-1
; Sequence 1, Application US/10329960
; Publication No. US2003009277A1
; GENERAL INFORMATION:
; APPLICANT: Fleischmann et al.
; TITLE OF INVENTION: Nucleotide Sequence of the Haemophilus influenzae Rd Genome, Frag
; TITLE OF INVENTION: Thereof, and Uses Thereof
; FILE REFERENCE: PB186P1
; CURRENT APPLICATION NUMBER: US/10/329,960
; CURRENT FILING DATE: 2003-01-02
; PRIOR APPLICATION NUMBER: US 09/643,990
; PRIOR FILING DATE: 2000-08-23
; PRIOR APPLICATION NUMBER: US 08/487,429
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: US 08/426,787
; PRIOR FILING DATE: 1995-04-21
; NUMBER OF SEQ ID NOS: 1
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 1830121
; TYPE: DNA
; ORGANISM: Haemophilus influenzae
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (4747)..(4747)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (9921)..(9921)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (10150)..(10150)
; OTHER INFORMATION: n equals a, t, g or c

; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (29298)..(29298)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (36543)..(36543)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (36551)..(36551)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (36636)..(36636)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (40808)..(40810)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (44416)..(44416)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (44905)..(44905)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (44975)..(44975)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (45593)..(45593)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (45732)..(45732)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (47036)..(47036)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (51334)..(51334)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (51602)..(51602)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (51786)..(51786)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (51805)..(51805)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (55369)..(55369)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (65309)..(65309)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (65313)..(65313)
; OTHER INFORMATION: n equals a, t, g or c
```



```
;
; NAME/KEY: misc_feature
; LOCATION: (80024)..(80024)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
;
; NAME/KEY: misc_feature
; LOCATION: (100091)..(100091)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
;
; NAME/KEY: misc_feature
; LOCATION: (102696)..(102696)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
;
; NAME/KEY: misc_feature
; LOCATION: (105121)..(105121)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
;
; NAME/KEY: misc_feature
; LOCATION: (107248)..(107248)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
;
; NAME/KEY: misc_feature
; LOCATION: (117136)..(117136)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
;
; NAME/KEY: misc_feature
; LOCATION: (119750)..(119750)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
;
; NAME/KEY: misc_feature
; LOCATION: (119924)..(119924)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
;
; NAME/KEY: misc_feature
; LOCATION: (120038)..(120038)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
;
; NAME/KEY: misc_feature
; LOCATION: (121344)..(121344)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
;
; NAME/KEY: misc_feature
; LOCATION: (122167)..(122167)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
;
; NAME/KEY: misc_feature
; LOCATION: (122336)..(122336)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
;
; NAME/KEY: misc_feature
; LOCATION: (131340)..(131340)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
;
; NAME/KEY: misc_feature
; LOCATION: (131360)..(131360)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
;
; NAME/KEY: misc_feature
; LOCATION: (139910)..(139910)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
;
; NAME/KEY: misc_feature
; LOCATION: (140398)..(140398)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
;
; NAME/KEY: misc_feature
; LOCATION: (142750)..(142750)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
;
; NAME/KEY: misc_feature
; LOCATION: (145058)..(145058)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
```

```
; NAME/KEY: misc_feature
; LOCATION: (145171)..(145171)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
;
; NAME/KEY: misc_feature
; LOCATION: (145942)..(145942)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
;
; NAME/KEY: misc_feature
; LOCATION: (147197)..(147197)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
;
; NAME/KEY: misc_feature
; LOCATION: (150841)..(150841)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
;
; NAME/KEY: misc_feature
; LOCATION: (152500)..(152500)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
;
; NAME/KEY: misc_feature
; LOCATION: (152530)..(152530)
;
;
Alignment Scores:
Pred. No.: 0.129 Length: 1830121
Score: 14.00 Matches: 14
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3.12% Indels: 0
DB: 14 Gaps: 0

US-09-701-229-2 (1-448) x US-10-329-960-1 (1-1830121)

Qy 157 GluteuTyrrValleuGluteuSerPheGlnLeuGluThr 170
Db 1203588 GAACCTTATGCTAGAGCTTCTAGTTTCAGCTTGAGACA 1203629

RESULT 7
US-09-815-242-3503/c
; Sequence 3503, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA 011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3503
; LENGTH: 135
```

```
; TYPE: DNA
; ORGANISM: Staphylococcus aureus
US-09-815-242-3503

Alignment Scores:
Pred. No.: 3 11 Length: 135
Score: 9.00 Matches: 9
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 2.01% Indels: 0
DB: 9 Gaps: 0

US-09-701-229-2 (1-448) x US-09-815-242-3503 (1-135)
QY 418 ValLeuLeuSerProAlaCysAlaSer 426
Db 107 GTATTATTGTCACCTGCTGTGCGAGT 81

RESULT 8
US-09-815-242-3597/c
; Sequence 3597, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011a
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3597
; LENGTH: 135
; TYPE: DNA
; ORGANISM: Staphylococcus aureus
US-09-815-242-3597

Alignment Scores:
Pred. No.: 3 11 Length: 135
Score: 9.00 Matches: 9
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 2.01% Indels: 0
DB: 9 Gaps: 0

US-09-701-229-2 (1-448) x US-09-815-242-3597 (1-135)
QY 418 ValLeuLeuSerProAlaCysAlaSer 426
Db 107 GTATTATTGTCACCTGCTGTGCGAGT 81

RESULT 9
US-09-815-242-1519/c
; Sequence 1519, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011a
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1519
; LENGTH: 216
; TYPE: DNA
; ORGANISM: Staphylococcus aureus
US-09-815-242-1519

Alignment Scores:
Pred. No.: 4 73 Length: 216
Score: 9.00 Matches: 9
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 2.01% Indels: 0
DB: 9 Gaps: 0

US-09-701-229-2 (1-448) x US-09-815-242-1519 (1-216)
QY 418 ValLeuLeuSerProAlaCysAlaSer 426
Db 34 GTATTATTGTCACCTGCTGTGCGAGT 8

RESULT 10
US-09-815-242-1487/c
; Sequence 1487, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011a
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
```

; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/269,308  
; PRIOR FILING DATE: 2001-02-16  
; NUMBER OF SEQ ID NOS: 14110  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 1487  
; LENGTH: 327  
; TYPE: DNA  
; ORGANISM: Staphylococcus aureus  
US-09-815-242-1487

Alignment Scores:  
Pred. No.: 6.85 Length: 327  
Score: 9.00 Matches: 9  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 2.01% Indels: 0  
DB: 9 Gaps: 0

US-09-701-229-2 (1-448) x US-09-815-242-1487 (1-327)

QY 418 ValLeuTeuSerProAlaCysAlaSer 426  
|||||  
DB 145 GTATTATTGTCACCTGCTGTGCGAGT 119

## RESULT 11

; Sequence 2986, Application US/09815242  
; Patent No. US20020061569A1  
; GENERAL INFORMATION:  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Kari L.  
; APPLICANT: Zyskind, Judith W.  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John D.  
; APPLICANT: Carr, Grant J.  
; APPLICANT: Yamamoto, Robert T.  
; APPLICANT: Xu, H. Howard  
; TITLE OF INVENTION: Identification of Essential Genes in  
; FILE REFERENCE: ELITRA.011A  
; CURRENT APPLICATION NUMBER: US/09/815,242  
; CURRENT FILING DATE: 2001-03-21  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/269,308  
; PRIOR FILING DATE: 2001-02-16  
; NUMBER OF SEQ ID NOS: 14110  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 2986  
; LENGTH: 383  
; TYPE: DNA  
; ORGANISM: Staphylococcus aureus  
US-09-815-242-2986

Alignment Scores:  
Pred. No.: 7.89 Length: 383  
Score: 9.00 Matches: 9  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 2.01% Indels: 0  
DB: 9 Gaps: 0

US-09-701-229-2 (1-448) x US-09-815-242-2986 (1-383)

QY 418 ValLeuTeuSerProAlaCysAlaSer 426  
|||||  
DB 39 GTATTATTGTCACCTGCTGTGCGAGT 13

## RESULT 12

US-10-027-632-180203/c  
; Sequence 180203, Application US/10027632  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
; FILE REFERENCE: 108827.129  
; CURRENT APPLICATION NUMBER: US/10/027,632  
; CURRENT FILING DATE: 2002-04-30  
; PRIOR APPLICATION NUMBER: US 60/218,006  
; PRIOR FILING DATE: 2000-07-12  
; PRIOR APPLICATION NUMBER: US 60/198,676  
; PRIOR FILING DATE: 2000-04-20  
; PRIOR APPLICATION NUMBER: US 60/193,483  
; PRIOR FILING DATE: 2000-03-29  
; PRIOR APPLICATION NUMBER: US 60/185,218  
; PRIOR FILING DATE: 2000-02-24  
; PRIOR APPLICATION NUMBER: US 60/167,363  
; PRIOR FILING DATE: 1999-11-23  
; PRIOR APPLICATION NUMBER: US 60/156,358  
; PRIOR FILING DATE: 1999-09-28  
; PRIOR APPLICATION NUMBER: US 60/146,002  
; PRIOR FILING DATE: 1999-08-09  
; NUMBER OF SEQ ID NOS: 325720  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 180203  
; LENGTH: 429  
; TYPE: DNA  
; ORGANISM: Human  
US-10-027-632-180203

Alignment Scores:  
Pred. No.: 8.73 Length: 429  
Score: 9.00 Matches: 9  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 2.01% Indels: 0  
DB: 13 Gaps: 0

US-09-701-229-2 (1-448) x US-10-027-632-180203 (1-429)

QY 312 TrpValArgGluArgGlnGlyValSer 320  
|||||  
DB 208 TGGGTGAGAGAAAGACAGGTTGTGAGT 182

## RESULT 13

US-10-027-632-180204/c  
; Sequence 180204, Application US/10027632  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
; FILE REFERENCE: 108827.129  
; CURRENT APPLICATION NUMBER: US/10/027,632  
; CURRENT FILING DATE: 2002-04-30  
; PRIOR APPLICATION NUMBER: US 60/218,006  
; PRIOR FILING DATE: 2000-07-12

;; PRIOR APPLICATION NUMBER: US 60/198,676  
;; PRIOR FILING DATE: 2000-04-20  
;; PRIOR APPLICATION NUMBER: US 60/193,483  
;; PRIOR FILING DATE: 2000-03-29  
;; PRIOR APPLICATION NUMBER: US 60/185,218  
;; PRIOR FILING DATE: 2000-02-24  
;; PRIOR APPLICATION NUMBER: US 60/167,363  
;; PRIOR FILING DATE: 1999-11-23  
;; PRIOR APPLICATION NUMBER: US 60/156,358  
;; PRIOR FILING DATE: 1999-09-28  
;; PRIOR APPLICATION NUMBER: US 60/146,002  
;; PRIOR FILING DATE: 1999-08-09  
;; NUMBER OF SEQ ID NOS: 325720  
;; SOFTWARE: FastSeq for Windows Version 4.0  
;; SEQ ID NO 180204  
;; LENGTH: 429  
;; TYPE: DNA  
;; ORGANISM: Human  
US-10-027-632-180204

Alignment Scores:  
Pred. No.: 8.73 Length: 429  
Score: 9.00 Matches: 9  
Percent Similarity: 100.00% Conservatives: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 2.01% Indels: 0  
DB: 13 Gaps: 0

US-09-701-229-2 (1-448) x US-10-027-632-180204 (1-429)

QY 312 TrpValArgGluArgGlnGlyValSer 320

DB 208 TGGGTGAGAGAAGACAGGGTGTGAGT 182

## RESULT 14

;; Sequence 2037, Application US/09815242  
;; Patent No. US20020061569A1  
;; GENERAL INFORMATION:  
;; APPLICANT: Haselbeck, Robert  
;; APPLICANT: Ohlsen, Kari L.  
;; APPLICANT: Zyskind, Judith W.  
;; APPLICANT: Wall, Daniel  
;; APPLICANT: Trawick, John D.  
;; APPLICANT: Carr, Grant J.  
;; APPLICANT: Yamamoto, Robert T.  
;; APPLICANT: Xu, H. Howard  
;; TITLE OF INVENTION: Identification of Essential Genes in  
;; FILE REFERENCE: ELITRA.011A  
;; CURRENT APPLICATION NUMBER: US/09/815,242  
;; CURRENT FILING DATE: 2001-03-21  
;; PRIOR APPLICATION NUMBER: 60/191,078  
;; PRIOR FILING DATE: 2000-03-21  
;; PRIOR APPLICATION NUMBER: 60/206,848  
;; PRIOR FILING DATE: 2000-05-23  
;; PRIOR APPLICATION NUMBER: 60/207,727  
;; PRIOR FILING DATE: 2000-05-26  
;; PRIOR APPLICATION NUMBER: 60/242,578  
;; PRIOR FILING DATE: 2000-10-23  
;; PRIOR APPLICATION NUMBER: 60/253,625  
;; PRIOR FILING DATE: 2000-11-27  
;; PRIOR APPLICATION NUMBER: 60/257,931  
;; PRIOR FILING DATE: 2000-12-22  
;; PRIOR APPLICATION NUMBER: 60/269,308  
;; PRIOR FILING DATE: 2001-02-16  
;; NUMBER OF SEQ ID NOS: 14110  
;; SOFTWARE: FastSeq for Windows Version 4.0  
;; SEQ ID NO 2037  
;; LENGTH: 485  
;; TYPE: DNA  
;; ORGANISM: Staphylococcus aureus  
US-09-815-242-2037

Alignment Scores:  
Pred. No.: 9.74 Length: 485  
Score: 9.00 Matches: 9  
Percent Similarity: 100.00% Conservatives: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 2.01% Indels: 0  
DB: 9 Gaps: 0

US-09-701-229-2 (1-448) x US-09-815-242-2037 (1-485)

QY 418 ValLeuLeuSerProAlaCysAlaSer 426

DB 287 GTATTATTGTACCTGCTGTGCGAGT 261

## RESULT 15

;; Sequence 280432, Application US/10027632  
;; GENERAL INFORMATION:  
;; APPLICANT: Wang, David G.  
;; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
;; FILE REFERENCE: 108827.129  
;; CURRENT APPLICATION NUMBER: US/10/027,632  
;; CURRENT FILING DATE: 2002-04-30  
;; PRIOR APPLICATION NUMBER: US 60/218,006  
;; PRIOR FILING DATE: 2000-07-12  
;; PRIOR APPLICATION NUMBER: US 60/198,676  
;; PRIOR FILING DATE: 2000-04-20  
;; PRIOR APPLICATION NUMBER: US 60/193,483  
;; PRIOR FILING DATE: 2000-03-29  
;; PRIOR APPLICATION NUMBER: US 60/185,218  
;; PRIOR FILING DATE: 2000-02-24  
;; PRIOR APPLICATION NUMBER: US 60/167,363  
;; PRIOR FILING DATE: 1999-11-23  
;; PRIOR APPLICATION NUMBER: US 60/156,358  
;; PRIOR FILING DATE: 1999-09-28  
;; PRIOR APPLICATION NUMBER: US 60/146,002  
;; PRIOR FILING DATE: 1999-08-09  
;; NUMBER OF SEQ ID NOS: 325720  
;; SOFTWARE: FastSeq for Windows Version 4.0  
;; SEQ ID NO 280432  
;; LENGTH: 498  
;; TYPE: DNA  
;; ORGANISM: Human  
US-10-027-632-280432

Alignment Scores:  
Pred. No.: 9.98 Length: 498  
Score: 9.00 Matches: 9  
Percent Similarity: 100.00% Conservatives: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 2.01% Indels: 0  
DB: 13 Gaps: 0

US-09-701-229-2 (1-448) x US-10-027-632-280432 (1-498)

QY 224 LeuIleAlaAspThrValProCysTrp 232

DB 347 TTGATCGCTGATCTGTTCCATGCTGG 373

Search completed: August 14, 2003, 11:30:57  
Job time : 918 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: August 14, 2003, 09:10:41 ; Search time 2291 Seconds  
(without alignments)  
4752.687 Million cell updates/sec

Title: US-09-701-229-2

Perfect score: 448

Sequence: 1 MSLIASDHFRIVVGLGKSGM.....MFKNFERGRLFAKAVEELA 448

Scoring table:

OLIGO  
Xgapop 60.0 , Xgapext 60.0  
Ygapop 60.0 , Ygapext 60.0  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 22781392 seqs, 12152238056 residues

Word size: 1

Total number of hits satisfying chosen parameters: 45546778

Minimum DB seq length: 25

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n model -DRV=xlp  
-Q/cgn2\_1/USPTO.spool/US09701229/runat\_09082003\_161237\_5033/app\_query.fasta\_1.647  
-DB=EST -Qfmt=fastap -SUFFIX=olip2n.rst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi -LIST=45  
-DOALIGN=200 -THR\_SCORE=quality -THR\_MIN=1 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc  
-NORM=ext -HEAPSIZ=500 -MINLEN=25 -MAXLEN=2000000000  
-USER=US09701229.ecgn\_1.1.4237 @runat\_09082003\_161237\_5033 -NCPU=6 -ICPU=3  
-NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -YGAPOP=60 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database :

EST:\*  
1: em\_estha:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estmu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_htc:\*  
9: gb\_esti:\*  
10: gb\_est2:\*  
11: gb\_hic:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: em\_gss\_hum:\*  
18: em\_gss\_inv:\*  
19: em\_gss\_pln:\*  
20: em\_gss\_vrt:\*  
21: em\_gss\_fun:\*  
22: em\_gss\_mam:\*  
23: em\_gss\_mus:\*  
24: em\_gss\_pro:\*  
25: em\_gss\_rod:\*  
26: em\_gss\_phq:\*  
27: em\_gss\_vri:\*  
28: gb\_gss1:\*

29: gb\_gss2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
C 1	200	44.6	1445	29	BZ568959	BZ568959 pacs2-164
C 2	186	41.5	892	29	BZ579116	BZ579116 msh2_6205
C 3	41	9.2	1171	29	BZ559806	BZ559806 pacs2-164
C 4	21	4.7	452	10	BE184674	BE184674 PM4-HT068
C 5	11	2.5	928	29	BZ567557	BZ567557 pacs2-164
C 6	9	2.0	199	6	AL814233	AL814233 Triticum
C 7	9	2.0	279	12	BM370877	BM370877 EBR004_SQ
C 8	9	2.0	309	14	CB605003	CB605003 3529_1_64
C 9	9	2.0	349	14	D39320	D39320 RICR3327A R
C 10	9	2.0	359	14	C19538	C19538 C19538 Rice
C 11	9	2.0	360	10	BB870897	BB870897 BB870897
C 12	9	2.0	366	14	CA622878	CA622878 wlin.pk01
C 13	9	2.0	366	14	CA692865	CA692865 wlm96.pk0
C 14	9	2.0	380	9	A1544865	A1544865 fb69b08.x
C 15	9	2.0	380	28	AZ060728	AZ060728 RPCI-23-4
C 16	9	2.0	392	10	BE403471	BE403471 WHE0429.H
C 17	9	2.0	392	12	BI781117	BI781117 EBma03_SQ
C 18	9	2.0	411	12	BI777969	BI777969 EBR007_SQ
C 19	9	2.0	424	9	AA510452	AA510452 VQ32106.I
C 20	9	2.0	429	9	AU162673	AU162673 AU162673
C 21	9	2.0	431	9	AW315553	AW315553 12985 MAR
C 22	9	2.0	442	9	AU183975	AU183975 AU183975
C 23	9	2.0	453	9	AW019231	AW019231 fd61a10.x
C 24	9	2.0	456	9	AU064507	AU064507 AU064507
C 25	9	2.0	472	9	AI667240	AI667240 fd14h09.x
C 26	9	2.0	477	10	BG605535	BG605535 WHE2239.B
C 27	9	2.0	484	10	BF200711	BF200711 WHE0821-0
C 28	9	2.0	487	12	BG907641	BG907641 LGL161F
C 29	9	2.0	497	9	AW564083	AW564083 LGL162.C
C 30	9	2.0	498	12	BM377980	BM377980 EBma04_SQ
C 31	9	2.0	499	9	AA136741	AA136741 ZK99A05.I
C 32	9	2.0	512	28	AZ098624	AZ098624 RPCI-23-2
C 33	9	2.0	515	12	BM323975	BM323975 PTC1_29.H
C 34	9	2.0	517	10	BE640499	BE640499 945004C03
C 35	9	2.0	531	28	AZ769198	AZ769198 LM0569M03
C 36	9	2.0	533	14	CB878729	CB878729 HP09316T
C 37	9	2.0	538	12	BM534621	BM534621 fx69e10.Y
C 38	9	2.0	540	14	CA700810	CA700810 wkmlc.pk0
C 39	9	2.0	549	12	BM859675	BM859675 fy59f07.x
C 40	9	2.0	559	10	BE510292	BE510292 946051F02
C 41	9	2.0	562	14	CA618126	CA618126 wlin.pk00
C 42	9	2.0	562	14	CB426323	CB426323 601498.WA
C 43	9	2.0	566	10	BG605315	BG605315 WHE2331.C
C 44	9	2.0	569	14	CA686112	CA686112 wlm96.pk0
C 45	9	2.0	579	13	BU371697	BU371697 603597031

ALIGNMENTS

RESULT 1  
BZ568959/c  
LOCUS  
DEFINITION  
BZ568959  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE

BZ568959  
pacs2-164\_8171.y2 pacs2-164  
Pseudomonas aeruginosa genomic clone  
pacs2-164\_8171, genomic survey sequence.  
BZ568959  
GI:27202799  
GSS.  
Pseudomonas aeruginosa  
Pseudomonas aeruginosa  
Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;  
Pseudomonadaceae; Pseudomonas.  
1 (bases 1 to 1445)

1445 bp  
DNA  
linear  
GSS 17-DEC-2002

AUTHORS Spencer, D.H., Raymond, C.K., Smith, E.E., Sims, E.E., Hastings, M., Burns, J.L., Kaul, R. and Olsen, M.V.  
 TITLE Whole-Genome-Sequence variation among multiple isolates of Pseudomonas aeruginosa library  
 JOURNAL J. Bacteriol., (2002) In press  
 COMMENT Contact: Chris K. Raymond  
 Genome Center  
 University of Washington  
 Box 352145, Seattle, WA 98105-2145, USA  
 Tel: 2062216954  
 Fax: 2066857244  
 Email: craymondeu.washington.edu  
 Class: shotgun.

FEATURES source  
 1. 1445  
 /organism="Pseudomonas aeruginosa"  
 /mol\_type="genomic DNA"  
 /strain="2-164"  
 /db\_xref="taxon:287"  
 /clone\_lib="pacs2-164\_8171"  
 /clone\_lib="pacs2-164"  
 /note="clinical isolate 2-164 Whole genomic shotgun library."

BASE COUNT 308 a 424 c 412 g 293 t 8 others  
 ORIGIN

Alignment Scores:  
 Pred. No.: 2.63e-179 Length: 1445  
 Score: 200.00 Matches: 200  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 44.64% Indels: 0  
 DB: 29 Gaps: 0

US-09-701-229-2 (1-448) x B2568959 (1-1445)

QY 147 ProAlaLeuAspLeuAlaAspAspIleGluLeuTyrValLeuGluLeuSerSerPhe 166  
 DB 675 CCGCGCTCGACCTCTCTGGCGACGACATCGAGCTGTACGTGTGGAGCTTCGAGCTTC 616

QY 167 GlnLeuGluThrCysAspArgLeuAsnAlaGluValAlaThrValLeuAsnValSerGlu 186  
 DB 615 CAGCTGGAAACCTGGATCGCTCAACCGCGAGGTGGCGACCGTGTCAAGCTCAGCGAA 556

QY 187 AspHisMetAspArgTyrAspGlyMetAlaAspTyrHisLeuAlaLysHisArgIlePhe 206  
 DB 555 GACCATATGGATCGCTACGCGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 496

QY 207 ArgGlyAlaArgGlnValValValValValValValValValValValValValValVal 226  
 DB 495 CGCGGTGCG 436

QY 227 AspThrValProCysTyrSerPheGlyLeuAsnLysProAspPheLysAlaPheGlyLeu 246  
 DB 435 GATACCGTGGCTGT 376

QY 247 IleGluGluAspGlyGlnLysTrpLeuAlaPheGlnPheAspLysLeuLeuProValGly 266  
 DB 375 ATCGAGGAAGACGGCCAGCAAGTGGCTGGCGTTCAGTTCGACAAAGCTGCTGCGGTTGGC 316

QY 267 GluLeuLysIleArgGlyAlaHisAsnTyrSerAsnAlaLeuAlaLeuAlaLeuGly 286  
 DB 315 GAACCTGAACATCGGTGGCGCCACCACTATTTCACACGGCGCTCGCGCGCTGGCGTGGC 256

QY 287 HisAlaValGlyLeuProPheAspAlaMetLeuGlyAlaLeuLysAlaPheSerGlyLeu 306  
 DB 255 CATCGGTGCGCTGCTGCGACGCTGCTGCGCGCTGCTGCGCGCTGCTGCGCGCTGCTGCG 196

QY 307 AlaHisArgCysGlnTrpValArgGlnGlyValSerTyrTyrAspAspSerLys 326  
 DB 195 GCTCATCGCTCCAGTGGGTACGGAGCGGCGGCGGTGAGCTACTACGACGATTTCGAAG 136

QY 327 AlaThrAsnValGlyAlaAlaLeuAlaAlaIleGluGlyLeuGlyAlaAspIleAspGly 346

DB 135 GCCACCAACGTCGGCGCGCTCTGGCGGATCGAGGGCTGGTGGCGATCGACACGGG 76

RESULT 2  
 B2579116/c  
 LOCUS msh2\_6205.y2 msh Pseudomonas aeruginosa genomic clone msh2\_6205,  
 DEFINITION genomic survey sequence.  
 ACCESSION B2579116  
 VERSION B2579116.1 GI:27214177  
 KEYWORDS GSS.  
 SOURCE Pseudomonas aeruginosa  
 ORGANISM Pseudomonas aeruginosa  
 Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;  
 Pseudomonadaceae; Pseudomonas.  
 REFERENCE 1 (bases 1 to 892)  
 AUTHORS Spencer, D.H., Raymond, C.K., Smith, E.E., Sims, E.E., Hastings, M., Burns, J.L., Kaul, R. and Olsen, M.V.  
 TITLE Whole-Genome-Sequence variation among multiple isolates of Pseudomonas aeruginosa library  
 JOURNAL J. Bacteriol., (2002) In press  
 COMMENT Contact: Chris K. Raymond  
 Genome Center  
 University of Washington  
 Box 352145, Seattle, WA 98105-2145, USA  
 Tel: 2062216954  
 Fax: 2066857244  
 Email: craymondeu.washington.edu  
 Class: shotgun.

FEATURES source  
 1. 892  
 /organism="Pseudomonas aeruginosa"  
 /mol\_type="genomic DNA"  
 /strain="MSH"  
 /db\_xref="taxon:287"  
 /clone\_lib="msh2\_6205"  
 /clone\_lib="msh"  
 /note="Environmental isolate. Whole genomic shotgun library."

BASE COUNT 173 a 277 c 276 g 160 t 6 others  
 ORIGIN

Alignment Scores:  
 Pred. No.: 3.16e-166 Length: 892  
 Score: 186.00 Matches: 186  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 41.52% Indels: 0  
 DB: 29 Gaps: 0

US-09-701-229-2 (1-448) x B2579116 (1-892)

QY 91 AlaLysGlyValArgIleSerGlyAspIleAspLeuPheAlaArgGluAlaLysAlaPro 110  
 DB 642 GCGAAAGCGTGGCATCTCCGTGACATCATCTTCGCCCGGAGCGGAGCGCCG 583

QY 111 IleValAlaIleThrGlySerAsnAlaLysSerThrValThrLeuValGlyGluMet 130  
 DB 582 ATCTGTCGCATCACCGGTTCCACGCGAAGACCGCTGACCCCTGGTGGGGAATG 523

QY 131 AlaValAlaAlaAspLysArgValAlaValGlyGlyAsnLeuGlyThrProAlaLeuThr 150  
 DB 522 CGGTGGCGCGGGAACAGCGTGTCCGCGTCCGCGGCAACCTCGGACCCCGGCTCGAC 463

QY 151 LeuLeuAlaAspAspIleGluLeuTyrValLeuGluLeuSerSerPheGlnLeuGluThr 170  
 DB 462 CTGCTGGCGGACGACATCGAGCTGTAGCTGTGTGGAGCTGTGAGCTTCCAGCTGGAAC 403

QY 171 CysAspArgLeuAsnAlaGluValAlaThrValLeuAsnValSerGluAspHisMetasp 190  
 DB 402 TGCATCGCTCAACGCGGAGGTGGCGACCGCTGCTGAACGCTCAGGAGGACCATATGAT 343

QY 191 ArgTyrAspGlyMetAlaAspTyrHisLeuAlaLysHisArgIlePheArgGlyAlaArg 210

```

Db      342  CGCTACGACGGCATGGCTGACTACCACTGCCAAGCACCGGATCTTCGCGGTGCCGC 283
QY      211  GlnValValAlaAsnArgAlaAspAlaLeuThrArgProLeuLeuAlaAspThrValPro 230
Db      282  CAGGTGCGTGTGAATCGCGCGGATGCCCTGACCCGACCGCGGTGATCGCCGATACCGTGGCC 223
QY      231  CysTrpSerPheGlyLeuAsnLysProAspPheLysAlaPheGlyLeuLeuGluAsp 250
Db      222  TGTGTGCTGTCGGCTGACACAGCGGACTTCAAGGCTTCGCGCTGATCGAGGAGAC 163
QY      251  GlyGlnLysTrpLeuAlaPheGlnPheAspLysLeuLeuProValGlyGluLeuLysIle 270
Db      162  GGCACAGAGTGGTGGCTGCTCCAGTTCGACAGCTGTCGCGGTGGCGAAGTCAAGATC 103
QY      271  ArgGlyAlaHisAsnTyr 276
Db      102  CGTGGCGGCCACCAACTAT 85

RESULT 3
LOCUS      BZ559806                      1171 bp      DNA      linear      GSS 17-DEC-2002
DEFINITION pacs2-164_1863.x1 pacs2-164 Pseudomonas aeruginosa genomic clone
ACCESSION  BZ559806
VERSION    BZ559806.1  GI:27176698
KEYWORDS   GSS.
SOURCE     Pseudomonas aeruginosa
ORGANISM   Pseudomonas aeruginosa
REFERENCE  1 (bases 1 to 1171)
AUTHORS    Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M.,
            Burns,J.L., Kaul,R. and Olsen,M.V.
TITLE      Whole-Genome-Sequence variation among multiple isolates of
            Pseudomonas aeruginosa library
JOURNAL    J. Bacteriol., (2002) In press
COMMENT    Contact: Chris K. Raymond
            Genome Center
            University of Washington
            Box 352145, Seattle, WA 98105-2145, USA
            Tel: 2062216954
            Fax: 2066857244
            Email: craymond@u.washington.edu
            Class: shotgun.

FEATURES   source
            1..1171
            /organism="Pseudomonas aeruginosa"
            /mol_type="genomic DNA"
            /strain="2-164"
            /db_xref="taxon:287"
            /clone="pacs2-164_1863"
            /clone_lib="pacs2-164"
            /note="clinical isolate 2-164 Whole genomic shotgun
            library."

BASE COUNT 218 a 325 c 329 g 298 t 1 others
ORIGIN

Alignment Scores:
Pred. No.:      3.47e-28      Length:      1171
Score:          41.00      Matches:      68
Percent Similarity: 97.14%      Conservative: 0
Best Local Similarity: 97.14%      Mismatches: 1
Query Match:    9.15%      Indels:      2
DB:              29      Gaps:      0

US-09-701-229-2 (1-448) x BZ559806 (1-1171)
QY      310  CysGlnTrpValArgGluArgGlnGlyValSerTyrTyrAspAspSerLysAlaThrAsn 329
Db      150  TGCAGTGGGTAGCGGAGCGGAGCGGTGAGTACTACGACGATCCCAAGGCCACCAAC 209
QY      330  ValGlyAlaAlaLeuAlaAlaIleGluGlyLeuGlyAlaAspGlyLysLeuVal 349

```

```

Db      210  GTCCGGCGCCCTCGCGGCGATCGAGGGCTGGTCCGACATCGACGCGCAAGCTGGTG 269
QY      350  LeuLeu-AlaGlyGlyAspGlyLysGlyAlaAspPheHisAspLeuArgGluProValAl 369
Db      270  CTGAT-CGCGCGGCGAGACGCGCAAGCGCGCGATTTTCATGACCTGCGGAGCGGTCGC 328
QY      369  aATGPhcCysArgAlaValValLeuLeu 378
Db      329  GCCTCTTTCGCGGCGGTGTACTGCTT 356

RESULT 4
LOCUS      BE184674                      452 bp      mRNA      linear      EST 22-JUN-2000
DEFINITION PM4-HT0688-050500-002-d10 HT0688 Homo sapiens cDNA, mRNA sequence.
ACCESSION  BE184674
VERSION    BE184674.1  GI:8663858
KEYWORDS   EST.
SOURCE     Homo sapiens (human)
ORGANISM   Homo sapiens
REFERENCE  1 (bases 1 to 452)
AUTHORS    Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordini,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.J.
TITLE      Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
JOURNAL    Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE    20202663
PUBMED     10737800
COMMENT    Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=et2-PM4-HT0688-050
            500-002-d10&t3=2000-05-05&t4=1)
            Seq primer: puc 18 forward
            High quality sequence start: 26
            High quality sequence stop: 89.

FEATURES   source
            1..452
            /organism="Homo sapiens"
            /mol_type="mRNA"
            /db_xref="taxon:9606"
            /dev_stage="Adult"
            /clone_lib="HT0688"
            /note="Organ: head neck; Vector: puc18; Site:1: SmaI;
            Site:2: SmaI; A mini-library was made by cloning products
            derived from ORESTES PCR (U.S. Letters patent application
            No. 196,716 - Ludwig Institute for Cancer Research)
            profiles into the pUC 18 vector. Reverse transcription of
            tissue mRNA and cDNA amplification were performed under
            low stringency conditions."

BASE COUNT 91 a 140 c 133 g 88 t
ORIGIN

Alignment Scores:
Pred. No.:      1.25e-09      Length:      452
Score:          21.00      Matches:      21
Percent Similarity: 100.00%      Conservative: 0
Best Local Similarity: 100.00%      Mismatches: 0
Query Match:    4.69%      Indels:      0
DB:              10      Gaps:      0

```



```

US-09-701-229-2 (1-448) x BE184674 (1-452)
QY 277 SerAsnAlaLeuAlaLeuAlaLeuGlyHisAlaValGlyLeuProPheAspAlaMet 296
    |||||
Db 241 TCCATGCGCTGGCGGCTTGGCACTCGGTACGCGGTGGCTGCCATTCGACCCCATG 300
QY 297 Leu 297
    |||
Db 301 CTC 303

RESULT 5
B2567557/c
LOCUS B2567557 928 bp DNA linear GSS 17-DEC-2002
DEFINITION pacs2-164_7094.y2 pacs2-164 Pseudomonas aeruginosa genomic clone
ACCESSION B2567557
VERSION B2567557
KEYWORDS B2567557.1 GI:27199005
SOURCE GSS.
ORGANISM Pseudomonas aeruginosa
REFERENCE 1 (bases 1 to 928)
AUTHORS Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M.,
        Burns,J.L., Kaul,R. and Olsen,M.V.
TITLE Whole-Genome-Sequence variation among multiple isolates of
        Pseudomonas aeruginosa library
JOURNAL J. Bacteriol., (2002) in press
COMMENT Contact: Chris K. Raymond
        Genome Center
        University of Washington
        Box 352145, Seattle, WA 98105-2145, USA
        Tel: 2062216954
        Fax: 2066857244
        Email: craymond@u.washington.edu
        Class: shotgun.

FEATURES             Location/Qualifiers
     source            1..928
                        /organism="Pseudomonas aeruginosa"
                        /mol_type="genomic DNA"
                        /strain="2-164"
                        /db_xref="taxon:287"
     clone="pacs2-164_7094"
     /clone_lib="pacs2-164"
     /note="clinical isolate 2-164 Whole genomic shotgun
        library."

BASE COUNT 238 a 228 c 236 g 226 t
ORIGIN
Alignment Scores:
Pred. No.: 9.12 Length: 928
Score: 11.00 Matches: 11
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 2.46% Indels: 0
DB: 29 Gaps: 0

US-09-701-229-2 (1-448) x B2567557 (1-928)
QY 279 AlaLeuAlaAlaLeuAlaLeuGlyHisAlaVal 289
    |||||
Db 324 CGGCTTGGCGCTGGCGGTGGCGCATCGGTC 292

RESULT 6
AL814293/c
ID AL814293 standard; RNA; EST; 199 BP.
XX
AC AL814293;
XX
SV AL814293.1
XX
DT 05-JUL-2002 (Rel. 72, Created)

US-09-701-229-2 (1-448) x AL814293 (1-199)
QY 348 LeuValLeuAlaGlyAspGly 356
    |||||
Db 88 CTCGTCTCTCTCGCGGTGGTGACGGC 62

RESULT 7
BM370877/c
LOCUS BM370877 279 bp mRNA linear EST 23-JUL-2002
DEFINITION EBR004_SQ002_K07_R root, 3 week, salt-stressed, cv Optic, EBR004
        Hordeum vulgare subsp. vulgare cDNA clone EBR004_SQ002_K07 5', mRNA
        sequence.
ACCESSION BM370877
VERSION BM370877.2 GI:21949180
KEYWORDS EST.
SOURCE Hordeum vulgare subsp. vulgare
ORGANISM Hordeum vulgare subsp. vulgare
REFERENCE 1 (bases 1 to 279)
AUTHORS Hedley,P., Liu,H., Caldwell,D., McCallum,N., Mudie,S., Cardie,L.,
        Ramsay,L., Machray,G., Marshall,D.F.M. and Waugh,R.
TITLE Development of Barley Transcriptome Resources
JOURNAL Unpublished

```

```

DT 02-SEP-2002 (Rel. 72, Last updated, Version 2)
XX Triticum aestivum (mercia) clone E12_e512_plate_7 maternal tissue (minus
DE endosperm and embryo) 8 days post anthesis
XX EST; expressed sequence tag.
XX Triticum aestivum (bread wheat)
OS Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae;
OC Triticeae; Triticum.
XX [1]
RN Wilson I., Beswick R., Shepherd S., Barker G., Parker J., Owen P.,
RA Edwards D., Coghill J., Holdsworth M., Lenton J., Shewry P., Edwards K.;
RT "A BBSRC-funded wheat EST resource for the academic community";
RL Unpublished.
XX [2]
RN Barker G.;
RP 1-199
RA Submitted (03-JUL-2002) to the EMBL/GenBank/DBJ databases.
RL Barker G., Institute of Arable Crop Research, Long Ashton, Bristol BS41 9AF
RL United Kingdom.
XX Key Location/Qualifiers
FH 1..199
FT /db_xref="taxon:4565"
FT /mol_type="mRNA"
FT /organism="Triticum aestivum"
FT /cultivar="mercia"
FT /dev_stage="8 days post anthesis"
FT /tissue_type="maternal tissue (minus endosperm and embryo)"
FT /clone_lib="e512"
FT /clone="E12_e512_plate_7"
XX
SQ Sequence 199 BP; 48 A; 56 C; 73 G; 22 T; 0 other;

Alignment Scores:
Pred. No.: 129 Length: 199
Score: 9.00 Matches: 9
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 2.01% Indels: 0
DB: 6 Gaps: 0

US-09-701-229-2 (1-448) x AL814293 (1-199)
QY 348 LeuValLeuAlaGlyAspGly 356
    |||||
Db 88 CTCGTCTCTCTCGCGGTGGTGACGGC 62

RESULT 7
BM370877/c
LOCUS BM370877 279 bp mRNA linear EST 23-JUL-2002
DEFINITION EBR004_SQ002_K07_R root, 3 week, salt-stressed, cv Optic, EBR004
        Hordeum vulgare subsp. vulgare cDNA clone EBR004_SQ002_K07 5', mRNA
        sequence.
ACCESSION BM370877
VERSION BM370877.2 GI:21949180
KEYWORDS EST.
SOURCE Hordeum vulgare subsp. vulgare
ORGANISM Hordeum vulgare subsp. vulgare
REFERENCE 1 (bases 1 to 279)
AUTHORS Hedley,P., Liu,H., Caldwell,D., McCallum,N., Mudie,S., Cardie,L.,
        Ramsay,L., Machray,G., Marshall,D.F.M. and Waugh,R.
TITLE Development of Barley Transcriptome Resources
JOURNAL Unpublished

```

## COMMENT

On Jan 10, 2002 this sequence version replaced gi:18114267.  
 Contact: Waugh R, Marshall DF  
 Genome Dynamics/Computational Biology  
 Scottish Crop Research Institute  
 Invergowrie, Dundee, DD2 5DA, Scotland, UK  
 Tel: 00 44 1382 562731  
 Fax: 00 44 1382 562426  
 Email: est@scri.sari.ac.uk  
 All sequence has a Phred quality score of 20 or over  
 Seq primer: M13 reverse.

## FEATURES

source  
 Location/Qualifiers

1. .279

/organism="Hordeum vulgare subsp. vulgare"

/mol\_type="mRNA"

/cultivar="Optic"

/db\_xref="taxon:112509"

/clone="EBro04\_SQ002\_K07"

/tissue\_type="root"

/dev\_stage="3 week"

/lab\_host="PH108"

/note="lib="root, 3 week, salt-stressed, cv Optic, Ebro04"  
 /clone="Vector: pSPORT1; Site.1: Sal I; Site.2: Not I;  
 Non-normalised library, directionally cloned into pSPORT1.  
 Derived from roots of 3 week old salt stressed barley  
 plants. Developed as part of the barley transcriptome  
 resources of BBSC/SEERAD funded cereal IGF (Investigating  
 Gene Function) project."

BASE COUNT 48 a 103 c 77 g 51 t

## ORIGIN

Alignment Scores:  
 Pred. No.: 189 Length: 279  
 Score: 9.00 Matches: 9  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 2.01% Indels: 0  
 DB: 12 Gaps: 0

US-09-701-229-2 (1-448) x BM370877 (1-279)

QY 397 ArgValAlaThrLeuAspGluAlaVal 405

Db 66 CGCGTCGACACCCCTTGACGAGCAGC 40

## RESULT 8

CB605003/c

LOCUS

DEFINITION CB605003 309 bp mRNA linear EST 16-MAY-2003

3529\_1\_64\_1\_E07.Y.1 3529 - 2 mm ear tissue from Schmidt and Hake

labs Zea mays cDNA, mRNA sequence.

ACCESSION CB605003.1 GI:29544623

VERSION EST.

KEYWORDS Zea mays

SOURCE Zea mays

## ORGANISM

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACAD

clade; Panicoideae; Andropogoneae; Zea.

1 (bases 1 to 309)

Walbot, V.

Maize ESTs from various cDNA libraries sequenced at Stanford

University

Unpublished

Contact: Walbot V

Department of Biological Sciences

Stanford University

855 California Ave, Palo Alto, CA 94304, USA

Tel: 650 723 2227

Fax: 650 725 8221

Email: walbot@stanford.edu

Plate: 3529.1\_64.1 row: E column: 07.

Location/Qualifiers

1. .309

/organism="Zea mays"

/mol\_type="mRNA"

/cultivar="B73"

/db\_xref="taxon:4577"

/tissue\_type="ear"

/dev\_stage="2 mm"

/lab\_host="E. coli XL0LR"

/clone\_lib="3529 - 2 mm ear tissue from Schmidt and Hake

labs"

/note="Organ: ear; Vector: PAD-GAL4-2.1; Site.1: EcoRI;

Site.2: XhoI; RNA isolated by Hake lab. 1 million pfu

amplified. Ampicillin is the selection marker."

BASE COUNT 72 a 88 c 113 g 36 t

## ORIGIN

Alignment Scores:  
 Pred. No.: 212 Length: 309  
 Score: 9.00 Matches: 9  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 2.01% Indels: 0  
 DB: 14 Gaps: 0

US-09-701-229-2 (1-448) x CB605003 (1-309)

QY 348 LeuValLeuAlaGlyGlyAspGly 356

Db 137 CTCGTCCTCTCGCGGTGTGACGCG 111

## RESULT 9

D39320/c

LOCUS

DEFINITION D39320 349 bp mRNA linear EST 03-APR-2002

RICR3327A Rice root Oryza sativa (Japonica cultivar-group) cDNA,

mRNA sequence.

ACCESSION D39320

VERSION D39320.1 GI:568471

KEYWORDS EST.

SOURCE Oryza sativa (Japonica cultivar-group)

Oryza sativa (Japonica cultivar-group)

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;

Ehrhartoidae; Oryzaceae; Oryza.

1 (bases 1 to 349)

Sasaki, T., Miyao, A. and Yamamoto, K.

Rice cDNA from callus 1995

Unpublished

Contact: Takuji Sasaki

National Institute of Agrobiological Resources

Rice Genome Research Program, Kannondai 2-1-2, Tsukuba, Ibaraki

305-8602, Japan

Tel: 81-298-38-7441

Fax: 81-298-38-7468

Email: tsasaki@abrr.affrc.go.jp, URL: http://rgp.dna.affrc.go.jp/

PROJECT "RGP"

Location/Qualifiers

1. .349

/organism="Oryza sativa (Japonica cultivar-group)"

/mol\_type="mRNA"

/db\_xref="taxon:39947"

/clone\_lib="Rice root"

/note="Prepared from seedling root."

BASE COUNT 72 a 110 c 118 g 44 t 5 others

## ORIGIN

Alignment Scores:  
 Pred. No.: 243 Length: 349  
 Score: 9.00 Matches: 9  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 2.01% Indels: 0  
 DB: 14 Gaps: 0

US-09-701-229-2 (1-448) x D39320 (1-349)

QY 348 LeuValLeuAlaGlyGlyAspGly 356  
 |||||||  
 Db 228 CTCGCTCCTCGCGGTGGTGCACGC 202

RESULT 10  
 C19538/c  
 LOCUS  
 DEFINITION C19538 Rice panicle at ripening stage Oryza sativa (japonica cultivar-group) cDNA clone E10575\_1A, mRNA sequence.

ACCESSION C19538  
 VERSION C19538.1 GI:1631809  
 KEYWORDS EST.  
 SOURCE Oryza sativa (japonica cultivar-group)  
 ORGANISM Oryza sativa (japonica cultivar-group)  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzaceae; Oryza.  
 1 (bases 1 to 359)

REFERENCE  
 AUTHORS Sasaki,T.  
 TITLE Rice cDNA from panicle at ripening stage  
 JOURNAL Unpublished  
 COMMENT Contact: Takuji Sasaki  
 National Institute of Agrobiological Resources  
 Rice Genome Research Program, Kannondai 2-1-2, Tsukuba, Ibaraki 305-8602, Japan  
 Tel: 81-298-38-7441  
 Fax: 81-298-38-7468  
 Email: tsasaki@abr.affrc.go.jp, URL:http://rgp.dna.affrc.go.jp/  
 PROJECT "RGP"

FEATURES  
 source  
 1..359  
 /organism="Oryza sativa (japonica cultivar-group)"  
 /mol\_type="mRNA"  
 /cultivar="Nipponbare"  
 /db\_xref="taxon:39947"  
 /clone="E10575\_1A"  
 /dev\_stage="ripening stage"  
 /clone\_lib="Rice panicle at ripening stage"  
 /note="Organ: panicle; Rice cDNA from panicle at ripening stage"

BASE COUNT 72 a 107 c 119 g 50 t 11 others

Alignment Scores:  
 Pred. No.: 251 Length: 359  
 Score: 9.00 Matches: 9  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 2.01% Indels: 0  
 DB: 14 Gaps: 0

US-09-701-229-2 (1-448) x C19538 (1-359)

QY 348 LeuValLeuAlaGlyGlyAspGly 356  
 |||||||  
 Db 174 CTCGCTCCTCGCGGTGGTGCACGC 148

RESULT 11  
 BB870897/c  
 LOCUS  
 DEFINITION BB870897 RIKEN full-length enriched, 10 days pregnant adult female ovary and uterus Mus musculus cDNA clone G630025M20 5', mRNA sequence.

ACCESSION BB870897  
 VERSION BB870897.1 GI:17117107  
 KEYWORDS EST.  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 360)

REFERENCE  
 AUTHORS Akimura,T., Arakawa,T., Carninci,P., Furuno,M., Hanagaki,T.,

Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T., Imotani,K., Ishii,Y., Ito,M., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Matsuyama,T., Nakamura,M., Nishi,K., Nomura,K., Numasaki,R., Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sakazume,N., Sasaki,D., Sato,K., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagawa,A., Takahashi,F., Takaku-Akahira,S., Tanaka,T., Tomaru,A., Toyota,T., Watahiki,A., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y.  
 RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura,T., et al. 2001)

Unpublished  
 Contact: Yoshihide Hayashizaki  
 Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute  
 The Institute of Physical and Chemical Research (RIKEN)  
 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
 Tel: 81-45-503-9222  
 Fax: 81-45-503-9216  
 Email: genome-res@gsc.riken.go.jp,  
 URL:http://genome.gsc.riken.go.jp/  
 Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.  
 Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)  
 wagi,K., Fujiwaka,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.  
 RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)  
 Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara,Y. and Hayashizaki,Y.  
 Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)  
 Please visit our web site (http://genome.gsc.riken.go.jp) for further details.  
 e mouse tissues.

FEATURES  
 source  
 1..360  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="G630025M20"  
 /sex="female"  
 /tissue\_type="ovary and uterus"  
 /dev\_stage="10 days pregnant adult"  
 /clone\_lib="RIKEN full-length enriched, 10 days pregnant adult female ovary and uterus"

BASE COUNT 96 a 112 c 85 g 67 t

ORIGIN

Alignment Scores:  
 Pred. No.: 252 Length: 360  
 Score: 9.00 Matches: 9  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 2.01% Indels: 0  
 DB: 10 Gaps: 0

US-09-701-229-2 (1-448) x BB870897 (1-360)

QY 348 LeuValLeuAlaGlyGlyAspGly 356  
 |||||||  
 Db 52 TTGGTGTCTCGGTGGTGGCGACGC 26

RESULT 12  
 CA622878/c  
 LOCUS  
 DEFINITION CA622878 wlln.pk0102.f2 wlln Triticum aestivum cDNA clone wlln.pk0102.f2 5' end, mRNA sequence.

ACCESSION CA622878  
 VERSION CA622878.1 GI:25201174  
 KEYWORDS EST  
 SOURCE Triticum aestivum (bread wheat)  
 ORGANISM Triticum aestivum

REFERENCE Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae ; Triticeae; Triticum.  
 1 (bases 1 to 366)  
 TINGEY,S.V., POWELL,W., WOLTERS,P., DOLAN,M., HAINEY,C., YUAN,Z., MAO,G., CARAHAN,N. and HANAFAY,M.K.  
 Authors DuPont Wheat cDNA Sequence  
 Title DuPont Wheat cDNA Sequence  
 Journal Unpublished  
 Comment Contact: Scott V. Tingey  
 Crop Genetics  
 E. I. DuPont de Nemours and Company  
 1 Innovation Way, P.O. Box 6104, Newark, DE 19714-6104, USA  
 Tel: 302-631-2602  
 Fax: 302-631-2607  
 Email: Scott.V.Tingey@USA.dupont.com  
 Seq primer: M13.

FEATURES  
 source Location/Qualifiers  
 1..366  
 /organism="Triticum aestivum"  
 /mol\_type="mrna"  
 /db\_xref="taxon:4565"  
 /clone="wlm96.pk0102.f2"  
 /tissue\_type="leaf"  
 /clone\_lib="wlm96"  
 /note="vector: pBluescript SK+; Site\_1: EcoRI; Site\_2: XhoI; Wheat (Triticum aestivum L.) leaf 7 day old seedling  
 XhoI; light grown (normalized)"  
 BASE COUNT 93 a 94 c 114 g 63 t 2 others  
 ORIGIN  
 Alignment Scores:  
 Pred. No.: 256 Length: 366  
 Score: 9.00 Matches: 9  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 2.01% Indels: 0  
 DB: 14 Gaps: 0

US-09-701-229-2 (1-448) x CA622878 (1-366)  
 QY 348 LeuValLeuAlaGlyGlyAspGly 356  
 |||||  
 Db 79 CTCGCTCTCTCGCCGGTGGGACGCC 53

RESULT 13  
 CA622878.1  
 LOCUS wlm96.pk060.b5 wlm96 Triticum aestivum cDNA clone wlm96.pk060.b5 5'  
 DEFINITION end, mRNA sequence.  
 VERSION CA622878.1 GI:2541444  
 KEYWORDS EST  
 SOURCE Triticum aestivum (bread wheat)  
 ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae ; Triticeae; Triticum.  
 1 (bases 1 to 366)  
 TINGEY,S.V., POWELL,W., WOLTERS,P., DOLAN,M., HAINEY,C., YUAN,Z., MAO,G., CARAHAN,N. and HANAFAY,M.K.  
 Authors DuPont Wheat cDNA Sequence  
 Title DuPont Wheat cDNA Sequence  
 Journal Unpublished  
 Comment Contact: Scott V. Tingey  
 Crop Genetics  
 E. I. DuPont de Nemours and Company  
 1 Innovation Way, P.O. Box 6104, Newark, DE 19714-6104, USA  
 Tel: 302-631-2602  
 Fax: 302-631-2607

Email: Scott.V.Tingey@USA.dupont.com  
 Seq primer: M13.  
 Location/Qualifiers  
 1..366  
 /organism="Triticum aestivum"  
 /mol\_type="mrna"  
 /cultivar="Stephens"  
 /db\_xref="taxon:4565"  
 /clone="wlm96.pk060.b5"  
 /tissue\_type="leaf"  
 /clone\_lib="wlm96"  
 /note="vector: pBluescript SK+; Site\_1: EcoRI; Site\_2: XhoI; Wheat (Triticum aestivum L.) seedlings 96 hr after inoculation w/ E. graminis"  
 BASE COUNT 65 a 125 c 114 g 53 t 9 others  
 ORIGIN  
 Alignment Scores:  
 Pred. No.: 256 Length: 366  
 Score: 9.00 Matches: 9  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 2.01% Indels: 0  
 DB: 14 Gaps: 0

US-09-701-229-2 (1-448) x CA692665 (1-366)  
 QY 348 LeuValLeuAlaGlyGlyAspGly 356  
 |||||  
 Db 266 CTCGCTCTCTCGCCGGTGGGACGCC 240

RESULT 14  
 AI544865/c  
 LOCUS AI544865 380 bp mRNA linear EST 07-JUN-2001  
 DEFINITION fb69b08.xl zebrafish WashU MPMG EST Danio rerio cDNA clone.  
 IMAGE:3717111 3', mRNA sequence.  
 ACCESSION AI544865  
 VERSION AI544865.1 GI:4462238  
 KEYWORDS EST  
 SOURCE Danio rerio (zebrafish)  
 ORGANISM Danio rerio  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes ; Cyprinidae; Danio.  
 REFERENCE 1 (bases 1 to 380)  
 AUTHORS Clark,M., Johnson,S.L., Lehrach,H., Lee,R., Li,F., Marra,M., Eddy,S., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R. and Wilson,R.  
 TITLE WashU zebrafish EST Project 1998  
 JOURNAL Unpublished  
 COMMENT Contact: Stephen L. Johnson  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: zbrafish@watson.wustl.edu  
 cDNA Library Preparation: Matthew Clark. cDNA Library Arrayed by: Matthew Clark. DNA Sequencing by: Washington University Genome Sequencing Center. Clone distribution: Genome Systems, St. Louis, Missouri (web address: www.genomesystems.com) (email contact: info@genomesystems.com) and Research Genetics, Huntsville, Alabama (web address: www.resgen.com) (email contact: info@resgen.com) and RessourcenzentrumPrimarDatenbank, Berlin, Germany (web address: www.rzpd.de)  
 Seg primer: T7 ET from Amersham  
 High quality sequence stop: 333  
 POLYA-NO. Location/Qualifiers  
 1..380  
 /organism="Danio rerio"

/mol\_type="mrna"  
/db\_xref="taxon:7955"  
/clone\_image="3717111"  
/sex="mixed"  
/tissue\_type="26 somite embryos, adult livers, shield  
stage embryos"  
/lab\_host="XLI-blue MRF"  
/clone\_lib="zebrafish WashU MPIMG EST"  
/note="Vector: pSPORT1; Site\_1: NotI; Site\_2: SalI; 1st  
strand cDNA was primed with a Not I - oligo(dT)15 primer  
[5']pGACTAGTCTAGATCGGAGCGGCCCTTTTCTTTT3';  
double-stranded cDNA was ligated to Sal I adaptors (BRL),  
digested with Not I and cloned into the Not I and Sal I  
sites of the pSPORT1 vector (BRL). Library was constructed  
by Matthew Clark (Lehrach lab; ICRF, London and Max Planck  
Institut fuer Molekulare Genetik, Berlin). cDNAs for EST  
analysis were selected following oligonucleotide  
hybridization fingerprinting of arrayed clones from  
zebrafish late somitogenesis (26 ss), adult liver or  
embryonic shield stage (5.6 h) libraries. Fingerprint  
data were used to computationally cluster cDNAs, and a  
single cDNA from each cluster was chosen for sequencing.  
In some cases multiple members of the same cluster were  
sequenced to assess clustering parameters or single clones  
were sequenced additional times to assess quality  
control."

BASE COUNT 119 a 78 c 82 g 101 t  
ORIGIN

Alignment Scores:  
Pred. No.: 267 Length: 380  
Score: 9.00 Matches: 9  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 2.01% Indels: 0  
DB: 9 Gaps: 0

US-09-701-229-2 (1-448) x A1544865 (1-380)

QY 78 GlyLeuSerLeuArgThProAlaLeu 86

DB 106 GGACTGAGTTTCGAGAACCCCTGCCCTA 80

RESULT 15

AZ060728/c

LOCUS

DEFINITION AZ060728 380 bp DNA linear GSS 30-MAR-2000

ACCESSION AZ060728

VERSION AZ060728.1 GI:7351899

KEYWORDS GSS.

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

1 (bases 1 to 380)

Zhao, S., Mierman, W., Feldblyum, T., Malek, J., Shatsman, S., Akinret

, B., Levins, M., McGann, S., Tsegaye, G., Geer, K., Krol, M., de Jong, P.

and Fraser, C.M.

Mouse BAC End Sequences from Library RPCI-23

Other\_GSSs: RPCI-23-415J24.TJ

Contact: Shaying Zhao

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 0200

Fax: 301 838 0208

Email: szhaoetigr.org

Clones are derived from the mouse BAC library RPCI-23. For BAC

library availability, please contact Pieter de Jong

(pieteredejong.med.buffalo.edu). Clones may be purchased from

BACPAC Resources (http://bacpac.med.buffalo.edu/orderingframe.htm)

or from Resea ch Genetics (info@resgen.com). BAC end page:  
http://www.tigr.org/tldb/bac\_ends/mouse/bac\_end\_intro.html  
Plate: 415 row: J column: 24

Seq primer: T7

Class: BAC ends.

FEATURES Location/Qualifiers

source

1..380

/organism="Mus musculus"

/mol\_type="genomic DNA"

/strain="C57BL/6J"

/db\_xref="taxon:10090"

/clone="RPCI-23-415J24"

/sex="Female"

/lab\_host="DH10B"

/clone\_lib="RPCI-23"

/note="Organ: Kidney/Brain; Vector: pBACE3.6; Site\_1:

EcoRI; Site\_2: EcoRI; Female C57BL/6J mouse kidney and/or

brain genomic DNA was isolated and partially digested

with a combination of EcoRI and EcoRI Methylase. Size

selected DNA was cloned into the pBACE3.6 vector at the

EcoRI sites. The ligation products were transformed into

DH10B electrocompetent cells (BRL Life Technologies).

BASE COUNT 132 a 71 c 88 g 89 t

ORIGIN

Alignment Scores:

Pred. No.: 267 Length: 380

Score: 9.00 Matches: 9

Percent Similarity: 100.00% Conservative: 0

Best Local Similarity: 100.00% Mismatches: 0

Query Match: 2.01% Indels: 0

DB: 28 Gaps: 0

US-09-701-229-2 (1-448) x AZ060728 (1-380)

QY 277 SerAsnAlaLeuAlaAlaLeuAlaLeu 285

DB 328 TCAAAATGCTTTAGCAGCTTGGCATTG 302

Search completed: August 14, 2003, 11:13:47

Job time : 2305 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

and is derived by analysis of the total score distribution.

SUMMARIES

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: August 14, 2003, 07:48:07 ; Search time 350 Seconds  
(without alignments)  
3455.281 Million cell updates/sec

Title: US-09-701-229-2

Perfect score: 448

Sequence: 1 MSLIASDHFRIVVGLKSGM.....MFKNFERGRLFAKAVEELA 448

Scoring table: OLIGO

Ygapop 60.0 , Xgapext 60.0

Ygapop 60.0 , Ygapext 60.0

Fgapop 6.0 , Fgapext 7.0

Delop 6.0 , Delext 7.0

Searched: 2552756 seqs, 1349719017 residues

Word size: 1

Total number of hits satisfying chosen parameters: 3293711

Minimum DB seq length: 25

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlp

-Q/Cp2\_1/USPTO.spool/US09701229/runat\_09082003\_161235\_5011/app\_query.fasta.1.647

-DB=N\_Geneseq\_19Jun03 -QFMT=fastap -SUFFIX=olip2n.rng -MINMATCH=0.1 -LOOPEL=0

-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=oligo -TRANS=human40.cdi

-LIST=45 -DOCALIGN=200 -THR\_SCORE=quality -THR\_MIN=1 -ALIGN=15 -MODE=LOCAL

-OUTFTW=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=25 -MAXLEN=2000000000

-USER=US09701229.ecgn\_1.1.471.e\_rnat\_09082003.161235\_5011 -NCPU=6 -ICPU=3

-NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG

-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6

-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database :

N\_Geneseq\_19Jun03:\*

1: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1980.DAT:\*

2: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1981.DAT:\*

3: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1982.DAT:\*

4: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1983.DAT:\*

5: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1984.DAT:\*

6: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1985.DAT:\*

7: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1986.DAT:\*

8: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1987.DAT:\*

9: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1988.DAT:\*

10: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1989.DAT:\*

11: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1990.DAT:\*

12: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1991.DAT:\*

13: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1992.DAT:\*

14: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1993.DAT:\*

15: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1994.DAT:\*

16: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1995.DAT:\*

17: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1996.DAT:\*

18: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1997.DAT:\*

19: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1998.DAT:\*

20: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1999.DAT:\*

21: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2000.DAT:\*

22: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT:\*

23: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT:\*

24: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2002.DAT:\*

25: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2003.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed,

Result No.	Score	Query Match	Length	DB ID	Description
1	448	100.0	1347	23	AAS54275 Pseudomonas aerugi
2	448	100.0	1450	21	AAZ47131 Pseudomonas aerugi
3	19	4.2	96109	22	AAF28548 Genomic fragment #
4	14	3.1	1080	23	AAS89258 DNA encoding novel
5	14	3.1	1080	23	AAS89801 DNA encoding novel
6	14	3.1	1080	23	AAS90166 DNA encoding novel
7	14	3.1	1314	22	AAS3453 Haemophilus influe
8	14	3.1	1317	22	AAH81460 Escherichia coli p
9	14	3.1	1317	23	AAS52295 E. coli DNA for ce
10	14	3.1	1317	23	AAS56341 Salmonella typhi D
11	14	3.1	2058	23	AAS73012 DNA encoding novel
12	14	3.1	3117	23	AAS60995 DNA encoding novel
13	14	3.1	3117	23	AAS89240 DNA encoding novel
14	14	3.1	4831	23	AAS94542 DNA encoding novel
15	14	3.1	4942	23	AAS89816 DNA encoding novel
16	14	3.1	4944	23	AAS89273 DNA encoding novel
17	14	3.1	4944	23	AAS90181 DNA encoding novel
18	14	3.1	1830121	17	AAT42063 Haemophilus influe
19	11	2.5	2703	23	AAS64579 DNA encoding novel
20	11	2.2	2703	23	AAS66661 DNA encoding novel
21	10	2.2	356	21	AAZ53173 Neisseria meningit
22	10	2.2	356	21	AAZ53174 Neisseria meningit
23	10	2.2	1335	25	ABZ41528 N. gonorrhoeae nuc
24	10	2.2	1380	21	AAZ53175 Neisseria meningit
25	10	2.2	69936	21	AAA81479 N. meningitidis pa
26	10	2.2	349980	21	AAF21607 Neisseria meningit
27	10	2.2	1437668	21	AAA81490 N. meningitidis B
28	9	2.0	135	23	AAS50926 Staphylococcus aur
29	9	2.0	135	23	AAS51020 Staphylococcus aur
30	9	2.0	216	23	AAS48942 Staphylococcus aur
31	9	2.0	285	21	AAZ53157 Neisseria gonorrh
32	9	2.0	285	21	AAZ53158 Neisseria meningit
33	9	2.0	285	21	AAZ53159 Neisseria meningit
34	9	2.0	297	25	ABX87811 Corn ear-derived p
35	9	2.0	298	25	ABX87421 Corn ear-derived p
36	9	2.0	327	23	AAS48910 Staphylococcus aur
37	9	2.0	383	23	AAS50409 Staphylococcus aur
38	9	2.0	485	23	AAS49460 Staphylococcus aur
39	9	2.0	717	18	AAH83927 DNA encoding a udp
40	9	2.0	717	19	AAV53378 DNA encoding a udp
41	9	2.0	1188	22	AAH52631 S. epidermidis ope
42	9	2.0	1242	20	AAX28684 Nucleotide sequenc
43	9	2.0	1320	18	AAV74391 Staphylococcus aur
44	9	2.0	1347	23	AAS51837 Staphylococcus aur
45	9	2.0	1350	23	AAS54613 Staphylococcus aur

ALIGNMENTS

RESULT 1
AAS54275
ID AAS54275 standard; DNA; 1347 BP.
AC AAS54275;
XX AAS54275;
DT 13-FEB-2002 (first entry)
XX Pseudomonas aeruginosa DNA for cellular proliferation protein #406.
DE Antisense; ds; prokaryotic cellular proliferation gene;
KW antibiotic; antibacterial; drug design.
XX Pseudomonas aeruginosa.
OS Pseudomonas aeruginosa.
XX WO200170955-A2.
PN XX

PD 27-SEP-2001:  
 XX 21-MAR-2001; 2001WO-US09180.  
 XX 21-MAR-2000; 2000US-191078P.  
 PR 23-MAY-2000; 2000US-206848P.  
 PR 26-MAY-2000; 2000US-207727P.  
 PR 23-OCT-2000; 2000US-242578P.  
 PR 27-NOV-2000; 2000US-253625P.  
 PR 22-DEC-2000; 2000US-257931P.  
 XX 16-FEB-2001; 2001US-269308P.  
 XX (ELIT-) ELITRA PHARM INC.  
 PA Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;  
 PI Yamamoto RT, Xu HH;  
 XX WPI; 2001-611495/70.  
 DR P-PSDB; AAU36416.  
 XX  
 PT New polynucleotides for the identification and development of  
 PT antibiotics, comprise sequences of antisense nucleic acids -  
 XX  
 PS Claim 27; Seq ID No 7912; 51pp; English.  
 CC The invention relates to antisense inhibitors of genes essential to  
 CC prokaryotic cellular proliferation, their use in identifying the  
 CC genes, their use in the discovery of novel antibiotics, the essential  
 CC genes themselves and the encoded proteins. The prokaryotes used are  
 CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella  
 CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The  
 CC invention is also useful for the identification of potential new targets  
 CC for antibiotic development. The antisense nucleic acids can also be used  
 CC to identify proteins used in proliferation, to express these proteins,  
 CC and to obtain antibodies capable of binding to the expressed proteins.  
 CC The proteins can be used to screen compounds in rational drug discovery  
 CC programmes. The antisense nucleic acid sequence is also useful to screen  
 CC for homologous nucleic acids which are required for cell proliferation in  
 CC a wide variety of organisms. The present sequence encodes an  
 CC essential prokaryotic cellular proliferation protein.  
 CC Note: The sequence data for this patent did not form part  
 CC of the printed specification, but was obtained in electronic  
 CC format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 1347 BP; 214 A; 453 C; 453 G; 227 T; 0 other;

Alignment Scores:  
 Pred. No.: 0 Length: 1347  
 Score: 448.00 Matches: 448  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 DB: 23 Gaps: 0

US-09-701-229-2 (1-448) x AAS54275 (1-1347)

Qy 1 MetSerLeuLeuAlaSerAspHisPheArgIleValValGlyLeuGlyLysSerGlyMet 20  
 Db 1 ATGAGCCTGATCGCTCCGACCACTTCGCGATCGTGTGCGGCTCGGCAAGCGGCATG 60  
 Qy 21 SerLeuValArgTyrLeuAlaArgGlyLeuProPheAlaValAspThrArgGlu 40  
 Db 61 TCCCTGTGTGCGCTACCTCGGCGCGCGGCTTTCGCGTGTGCGATACCCGAGAG 120  
 Qy 41 AsnProGluLeuAlaThrLeuArgAlaGlnTyrProGlnValValArgCysGly 60  
 Db 121 AACCCGCGGAGCTGGCCACCTCCGTCGCCAGTATCCGAGGTGGAGTGGTTCGGC 180  
 Qy 61 GluLeuAspAlaGluPheLeuCysSerAlaArgGluLeuTyrValSerProGlyLeuSer 80  
 Db 181 GAACCTGACGCGGAGTTCCTCTGCTCCGCGCGCAACTCTATGTCAGCCCGGCTTGTG 240

Qy 81 LeuArgThrProAlaLeuValGlnAlaAlaLysGlyValArgIleSerGlyAspIle 100  
 Db 241 CTGCGACCCCTGCTGCTACAGCCGCGGAAAGCGTGCATCTCCGGTGACATC 300  
 Qy 101 AspLeuPheAlaArgGluAlaLysAlaProIleValAlaIleThrGlySerAsnAlaLys 120  
 Db 301 GATCTCTTCGCGCGGAGCGGAAAGCGCGATCGTGCCTATCACCGGTTCACACGCGAAG 360  
 Qy 121 SerThrValThrThrLeuValGlyGluMetalValAlaAlaAspLysArgValAlaVal 140  
 Db 361 AGCACCGTGACCACTGCTGGCGGAATGGCGGTGGCGGACAGCGTGTCCGCCGTC 420  
 Qy 141 GlyGlyAsnLeuGlyThrProAlaLeuAspLeuLeuAlaAspAspIleGluLeuTyrVal 160  
 Db 421 GCGGCAACCTCGGCACCCCGCGCTCGACCTGCTGGCCGACGATCGAGTGTACGTG 480  
 Qy 161 LeuGluLeuSerSerPheGlnLeuGluThrCysAspArgLeuAsnAlaGluValAlaThr 180  
 Db 481 TTGGAGCTGTGAGCTTCCAGCTGGAAACCTGCGATCGCTCAACCGCGAGGTGGCGAC 540  
 Qy 181 ValLeuAsnValSerGluAspHisMetAspArgTyrAspGlyMetAlaAspTyrHisLeu 200  
 Db 541 GTGCTGAACGTCAGCGAAGACCATATGGATCGTACGACGCGCATGGCTGACTTACCACCTG 600  
 Qy 201 AlaLysHisArgIlePheArgGlyAlaArgGlnValValAlaAsnArgAlaAspAlaLeu 220  
 Db 601 GCCAAGCACCGGATCTTCGCGGTGGCGCCGAGGTGCTGGTGAATTCGCGCGCATGCCCTG 660  
 Qy 221 ThrArgProLeuIleAlaAspThrValProCysTyrTrpSerPheGlyLeuAsnLysProAsp 240  
 Db 661 ACCGACCGCTGATGCGCGATACCGTGGCTGCTGCTGCTGGCTGCTGCGCTGAACAGCGGAC 720  
 Qy 241 PheLysAlaPheGlyLeuIleGluAspGlyGlnLysTrpLeuAlaPheGlnPheAsp 260  
 Db 721 TTCAAGGCTTTCGGCTGATCGAGGAGACGCGCAGAGTGGCTGGCTTCCAGTTCGAC 780  
 Qy 261 LysLeuLeuProValGlyLeuLysIleArgGlyAlaHisAsnTyrSerAsnAlaLeu 280  
 Db 781 AAGTGTCTCGCGTGGCGAATCGATCGTGGCGCCCACTATTTCCAAACGCGCTC 840  
 Qy 281 AlaAlaLeuAlaLeuGlyHisAlaValGlyLeuProPheAspAlaMetLeuGlyAlaLeu 300  
 Db 841 GCCGCGTGGCGTGGCGCATGCGGTGCGCTGCGCTTCGACGCCATGCTCGGCGCGCTG 900  
 Qy 301 LysAlaPheSerGlyLeuAlaHisArgCysGlnTrpValArgGluArgGlnGlyValSer 320  
 Db 901 AAGCGTTCGCGCTGCTCATCGCTGCCAGTGGGTACGCGAGCGCGAGCGGCCGAT 960  
 Qy 321 TyrTyrAspAspSerLysAlaThrAsnValGlyAlaAlaLeuAlaAlaIleGluGlyLeu 340  
 Db 961 TACTACGACGATTCGAAGGCCACCAACGTCGCGCGCGCTTCCGCGCGGCTGCTGCTG 1020  
 Qy 341 GlyAlaAspIleAspGlyLysLeuValLeuAlaGlyLysArgGlyLysGlyAlaAsp 360  
 Db 1021 GGTCCGACATCGAGCGCACTGCTGCTGCTGCGCGCGAGAGCGGCAAGCGGCCGAT 1080  
 Qy 361 PheHisAspLeuArgGluProValAlaAlaArgPheCysArgAlaValAlaLeuGlyArg 380  
 Db 1081 TTCCATGACCTCGCGAGCGCGTTCGCGCGCTTCTCGCGCGGCTGGTACTGCTGGCGCT 1140  
 Qy 381 AspAlaGlyLeuIleAlaGlnAlaLeuGlyAsnAlaValProLeuValArgValAlaThr 400  
 Db 1141 GACCGCGGCTGATTTGCCAGGCACTGGGCAACGCGTACCGCTGGTGGCGCGCGCAACG 1200  
 Qy 401 LeuAspGluAlaValArgGlnAlaAlaGluLeuAlaArgGluGlyAspAlaValLeuLeu 420  
 Db 1201 CTGACCGAAGCAGTCCCGAGCGCGCTGCGCTGCGCGCGGAGCGGATGCGGCTGTG 1260  
 Qy 421 SerProAlaCysAlaSerLeuAspMetPheLysAsnPheGluGluArgGlyArgLeuPhe 440  
 Db 1261 TCGCGCGCTGCGCGAGCCTGGACATGTTCAAGAACTTCGAAGAACGCGGACGCGCTGTC 1320  
 Qy 441 AlaLysAlaValGluGluLeuAla 448



-----

QY 421 SerProAlaCyAlaSerLeuAspMetPheLysAsnPheGluGluArgGlyArgLeuPhe 440  
|||||  
Db 1311 TCGCGGGCTGGCGGAGCTGGACATGTTCAAGAACTTCGAAGAACGGGAGCGCTGTTC 1370  
|||  
QY 441 AlaLysAlaValGluGluLeuAla 448  
|||||  
Db 1371 GCCAAGCGGTAGAGGAGTAGCG 1394  
|||

RESULT 3  
AAF28548/C  
ID AAF28548 standard; DNA; 96109 BP.

XX AAF28548;

XX 04-APR-2001 (first entry)

XX Genomic fragment #35.

XX Genomic library; bacteria; human upper airway; otitis media; sinusitis;  
KW bronchopulmonary; endocarditis; meningitis; ss.

XX Moraxella catarrhalis.

XX WO200078968-A2.

XX 28-DEC-2000.

XX 16-JUN-2000; 2000WO-US16649.

XX 18-JUN-1999; 99US-0140121.

XX (INCY-) INCYTE GENOMICS INC.

XX Lagace RE, Patterson C, Berg KL;

XX WPI; 2001-041427/05.

XX Genomic library for identifying diagnostic and therapeutic  
PT compositions, and for identifying virulence factors, regulatory  
PT elements and drug targets, comprises Moraxella catarrhalis nucleic  
PT acids -

PS Claim 1; Page 345-368; 545pp; English.

XX The present invention relates to a Moraxella catarrhalis genomic library  
CC comprising of a combination of 41 nucleic acid molecules (see  
CC AAF28514-AAF28554). The library has a number of uses described in the  
CC specification e.g. is useful for identifying diagnostic and therapeutic  
CC compositions. M. catarrhalis (Branhamella catarrhalis) is a large  
CC aerobic, gram-negative diplococcus, normally found among the bacterial  
CC flora of human upper airways. M. catarrhalis is known to cause acute,  
CC localised infections such as otitis media, sinusitis and bronchopulmonary  
CC infection and life-threatening, systemic diseases including endocarditis  
CC and meningitis.

XX Sequence 96109 BP; 28783 A; 18910 C; 20341 G; 28075 T; 0 other;

Alignment Scores:

Pred. No.: 1.02e-06 Length: 96109  
Score: 19.00 Matches: 19  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 4.24% Indels: 0  
DB: 22 Gaps: 0

US-09-701-229-2 (1-448) x AAF28548 (1-96109)

QY 113 AlalleThrGlySerAsnAlaLysSerThrValThrThrLeuValGlyGluMetAla 131  
|||||

Db 12400 GCCATCAGCGCTCAATGCCAAAGACACTGTTACGACATTAGTAGGCGAGATGCCA 12344  
|||||

RESULT 4

AAS89258

ID AAS89258 standard; cDNA; 1080 BP.

XX AAS89258;

XX 13-FEB-2002 (first entry)

XX DNA encoding novel human diagnostic protein #25062.

XX Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.  
XX Homo sapiens.

XX WO200175067-A2.

XX 11-OCT-2001.

XX 30-MAR-2001; 2001WO-US08631.

XX 31-MAR-2000; 2000US-0540217.

XX 23-AUG-2000; 2000US-0649167.

XX (HYSE-) HYSEQ INC.

XX Drmanac RT, Liu C, Tang YT;

XX WPI; 2001-639362/73.

XX P-PSDB; ABG25071.

XX New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity

XX Claim 1; SEQ ID No 25062; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. AAS64197-AAS94564 represent novel human  
CC diagnostic coding sequences of the invention.

CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 1080 BP; 267 A; 253 C; 299 G; 261 T; 0 other;

Alignment Scores:

Pred. No.: 0.00104 Length: 1080  
Score: 14.00 Matches: 14  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 3.12% Indels: 0  
DB: 23 Gaps: 0

US-09-701-229-2 (1-448) x AAS89258 (1-1080)

QY 157 GluLeuTyrValLeuGluLeuSerSerPheGlnLeuGluThr 170  
|||||

Db 22 GAACGTACGTGCTGGAACTGTGAGCTTCAGCTCCAGCTGGAACC 63

RESULT 5  
AAS89801  
ID AAS89801 standard; cDNA; 1080 BP.  
XX  
AC AAS89801;  
XX  
DT 13-FEB-2002 (first entry)  
XX  
DE DNA encoding novel human diagnostic protein #25605.  
XX  
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO200175067-A2.  
XX  
PD 11-OCT-2001.  
XX  
PF 30-MAR-2001; 2001WO-US08631.  
XX  
PR 31-MAR-2000; 2000US-0540217.  
PR 23-AUG-2000; 2000US-0649167.  
XX  
PA (HYSE-) HYSEQ INC.  
XX  
PI Drmanac RT, Liu C, Tang YT;  
XX  
DR WPI; 2001-639362/73.  
DR P-PSDB; ABG25614.  
XX  
XX New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity  
XX  
XX

Claim 1; SEQ ID No 25605; 103pp; English.  
The invention relates to isolated polynucleotide (I) and  
polypeptide (II) sequences. (I) is useful as hybridisation probes,  
polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
and gene mapping, and in recombinant production of (II). The  
polynucleotides are also used in diagnostics as expressed sequence tags  
for identifying expressed genes. (I) is useful in gene therapy techniques  
to restore normal activity of (II) or to treat disease states involving  
(II). (II) is useful for generating antibodies against it, detecting or  
quantitating a polypeptide in tissue, as molecular weight markers and as  
a food supplement. (II) and its binding partners are useful in medical  
imaging of sites expressing (II). (I) and (II) are useful for treating  
disorders involving aberrant protein expression or biological activity.  
The polypeptide and polynucleotide sequences have applications in  
diagnostics, forensics, gene mapping, identification of mutations  
responsible for genetic disorders or other traits to assess biodiversity  
and to produce other types of data and products dependent on DNA and  
amino acid sequences. AAS64197-AAS94564 represent novel human  
diagnostic coding sequences of the invention.  
Note: The sequence data for this patent did not appear in the printed  
specification, but was obtained in electronic format directly from WIPO  
at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 1080 BP; 267 A; 253 C; 299 G; 261 T; 0 other;

Alignment Scores:  
Pred. No.: 0.00104 Length: 1080  
Score: 14.00 Matches: 14  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 3.12% Indels: 0  
Gaps: 0

US-09-701-229-2 (1-448) x AAS89801 (1-1080)

Oy 157 GluLeuTyrrValLeuGluLeuSerSerPheGlnLeuGluThr 170  
|||||  
Db 22 GAACGTACGTGCTGGAACTGTGAGCTTCAGCTGGAACC 63

RESULT 6  
AAS90166  
ID AAS90166 standard; cDNA; 1080 BP.  
XX  
AC AAS90166;  
XX  
DT 13-FEB-2002 (first entry)  
XX  
DE DNA encoding novel human diagnostic protein #25970.  
XX  
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO200175067-A2.  
XX  
PD 11-OCT-2001.  
XX  
PF 30-MAR-2001; 2001WO-US08631.  
XX  
PR 31-MAR-2000; 2000US-0540217.  
PR 23-AUG-2000; 2000US-0649167.  
XX  
PA (HYSE-) HYSEQ INC.  
XX  
PI Drmanac RT, Liu C, Tang YT;  
XX  
DR WPI; 2001-639362/73.  
DR P-PSDB; ABG25979.  
XX  
XX New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity  
XX  
XX

Claim 1; SEQ ID No 25970; 103pp; English.  
The invention relates to isolated polynucleotide (I) and  
polypeptide (II) sequences. (I) is useful as hybridisation probes,  
polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
and gene mapping, and in recombinant production of (II). The  
polynucleotides are also used in diagnostics as expressed sequence tags  
for identifying expressed genes. (I) is useful in gene therapy techniques  
to restore normal activity of (II) or to treat disease states involving  
(II). (II) is useful for generating antibodies against it, detecting or  
quantitating a polypeptide in tissue, as molecular weight markers and as  
a food supplement. (II) and its binding partners are useful in medical  
imaging of sites expressing (II). (I) and (II) are useful for treating  
disorders involving aberrant protein expression or biological activity.  
The polypeptide and polynucleotide sequences have applications in  
diagnostics, forensics, gene mapping, identification of mutations  
responsible for genetic disorders or other traits to assess biodiversity  
and to produce other types of data and products dependent on DNA and  
amino acid sequences. AAS64197-AAS94564 represent novel human  
diagnostic coding sequences of the invention.  
Note: The sequence data for this patent did not appear in the printed  
specification, but was obtained in electronic format directly from WIPO  
at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 1080 BP; 267 A; 253 C; 299 G; 261 T; 0 other;

Alignment Scores:  
Pred. No.: 0.00104 Length: 1080  
Score: 14.00 Matches: 14  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 3.12% Indels: 0

DB: 23 Gaps: 0

US-09-701-229-2 (1-448) x AAS90166 (1-1080)

QY 157 GluLeuTyrValLeuGluLeuSerSerPheGlnLeuGluThr 170  
|||||  
Db 22 GAACGTGACGTGGAAGCTGTCGAGCTCCAGCTGGAACC 63

RESULT 7  
AAS53453  
ID AAS53453 standard; DNA; 1314 BP.  
AC AAS53453;  
XX  
XX 13-FEB-2002 (first entry)  
DE Haemophilus influenzae DNA for cellular proliferation protein #235.  
XX  
XX Antisense; ds; prokaryotic cellular proliferation gene;  
KW antibiotic; antibacterial; drug design.  
XX  
XX Haemophilus influenzae.  
OS  
PN WO200170955-A2.  
XX  
XX 27-SEP-2001.  
PD  
XX  
XX 21-MAR-2001; 2001WO-US09180.  
XX  
XX 21-MAR-2000; 2000US-191078P.  
PR  
XX 23-MAY-2000; 2000US-206848P.  
PR  
XX 26-MAY-2000; 2000US-207727P.  
PR  
XX 23-OCT-2000; 2000US-242578P.  
PR  
XX 27-NOV-2000; 2000US-253625P.  
PR  
XX 22-DEC-2000; 2000US-257931P.  
PR  
XX 16-FEB-2001; 2001US-269308P.  
XX  
XX (ELIT-) ELITRA PHARM INC.  
XX  
XX Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;  
PI Yamamoto RT, Xu HH;  
PI  
XX  
XX WPI; 2001-611495/70.  
DR P-PSDB; AAS35594.  
XX  
XX New polynucleotides for the identification and development of  
PT antibiotics, comprise sequences of antisense nucleic acids -  
XX  
XX  
PS Claim 27; Seq ID No 7090; 51lpp; English.  
XX  
XX The invention relates to antisense inhibitors of genes essential to  
CC prokaryotic cellular proliferation, their use in identifying the  
CC genes, their use in the discovery of novel antibiotics, the essential  
CC genes themselves and the encoded proteins. The prokaryotes used are  
CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella  
CC pneumoniae, pseudomonas aeruginosa and Enterococcus faecalis. The  
CC invention is also useful for the identification of potential new targets  
CC for antibiotic development. The antisense nucleic acids can also be used  
CC to identify proteins used in proliferation, to express these proteins,  
CC and to obtain antibodies capable of binding to the expressed proteins.  
CC The proteins can be used to screen compounds in rational drug discovery  
CC programmes. The antisense nucleic acid sequence is also useful to screen  
CC for homologous nucleic acids which are required for cell proliferation in  
CC a wide variety of organisms. The present sequence encodes an  
CC essential prokaryotic cellular proliferation protein.  
CC Note: The sequence data for this patent did not form part  
CC of the printed specification, but was obtained in electronic  
CC format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 1314 BP; 403 A; 214 C; 294 G; 403 T; 0 other;

Alignment Scores:

Pred. No.: 0.00125 Length: 1314  
Score: 14.00 Matches: 14  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 3.12% Indels: 0  
DB: 23 Gaps: 0

US-09-701-229-2 (1-448) x AAS53453 (1-1314)

QY 157 GluLeuTyrValLeuGluLeuSerSerPheGlnLeuGluThr 170  
|||||  
Db 457 GAACTTTATGTACTAGAGCTTTCTAGTTTTCAGCTTGAGACA 498

RESULT 8  
AAH81460  
ID AAH81460 standard; DNA; 1317 BP.  
AC AAH81460;  
XX  
XX 21-SEP-2001 (first entry)  
DT  
XX  
XX Escherichia coli protein encoding nucleotide sequence SEQ ID NO:259.  
DE  
XX  
XX Escherichia coli; identification; proliferation; microorganism;  
KW antimicrobial; antibacterial; antibiotic; gene therapy; diagnosis;  
KW bacterial growth inhibition; ds.  
XX  
XX Escherichia coli.  
OS  
XX WO200148209-A2.  
PN  
XX 05-JUL-2001.  
PD  
XX 19-DEC-2000; 2000WO-US34419.  
PF  
XX 23-DEC-1999; 99US-0173005.  
PR  
XX (ELIT-) ELITRA PHARM INC.  
PA  
XX  
XX Forsyth RA, Ohlsen KL, Zyskind JW;  
PI  
XX WPI; 2001-457376/49.  
DR P-PSDB; AAG98404.  
XX  
XX Novel nucleic acids encoding proteins required for Escherichia coli  
PT proliferation, useful for screening for antimicrobial agents -  
XX  
XX Claim 9; Page 388-389; 596pp; English.  
PS  
XX  
XX The present invention describes a purified or isolated nucleic acid  
CC sequence (I) consisting essentially of one of the 93 nucleotide sequences  
CC given in AAH81202 to AAH81294, where expression of the nucleic acid in a  
CC microorganism is capable of inhibiting proliferation of a microorganism.  
CC (I) have antibacterial and antibiotic activities, and can be used in  
CC gene therapy. Expression of (I) in a microorganism inhibits proliferation  
CC of the microorganism, and the manufactured antibiotic is useful for  
CC reducing the activity or level of a gene product required for  
CC proliferation of a microorganism in a subject, specifically humans. The  
CC nucleic acids that inhibit bacterial growth or proliferation can be used  
CC as antisense therapeutics for killing bacteria. In addition to  
CC therapeutic applications, the nucleic acid sequences complementary to  
CC sequences required for proliferation can be used as diagnostic tools.  
CC For example, nucleic acid probes complementary to proliferation-required  
CC sequences that are specific for particular species of microorganisms can  
CC be used as probes to identify particular microorganism species in  
CC clinical specimens. AAH81295 to AAH81487 encode the Escherichia coli  
CC proteins given in AAG98239 to AAG98431, and AAH81488 to AAH81491  
CC represent oligonucleotides, which are used in the exemplification of the  
CC present invention.  
XX  
SQ Sequence 1317 BP; 294 A; 323 C; 400 G; 300 T; 0 other;

Alignment Scores:

Pred. No.: 0.00126 Length: 1317  
 Score: 14.00 Matches: 14  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 3.12% Indels: 0  
 DB: 22 Gaps: 0

US-09-701-229-2 (1-448) x AAH81460 (1-1317)

QY 157 GluLeuTyRValLeuGluLeuSerSerPheGlnLeuGluThr 170  
 |||||  
 Db 457 GAACTGTACGTGCTGGAACTGTCGAGCTTCCAGCTGGAAACC 498

## RESULT 9

AAS52295  
 ID AAS52295 standard; DNA; 1317 BP.  
 XX  
 AC AAS52295;  
 XX  
 DT 13-FEB-2002 (first entry)  
 XX  
 DE E. coli DNA for cellular proliferation protein #17.  
 XX  
 KW Antisense; ds; prokaryotic cellular proliferation gene;  
 KW antibiotic; antibacterial; drug design.  
 XX  
 OS Escherichia coli.  
 XX  
 PN WO200170955-A2.  
 XX  
 PD 27-SEP-2001.  
 XX  
 PF 21-MAR-2001; 2001WO-US09180.  
 XX  
 PR 21-MAR-2000; 2000US-191078P.  
 PR 23-MAY-2000; 2000US-206848P.  
 PR 26-MAY-2000; 2000US-207727P.  
 PR 23-OCT-2000; 2000US-242578P.  
 PR 27-NOV-2000; 2000US-253625P.  
 PR 22-DEC-2000; 2000US-257931P.  
 PR 16-FEB-2001; 2001US-269308P.  
 XX  
 PA (ELIT-) ELITRA PHARM INC.

Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;  
 Yamamoto RT, Xu HH;  
 WPI: 2001-611495/70.  
 P-PSDB; AAU34436.  
 New polynucleotides for the identification and development of  
 antibiotics, comprise sequences of antisense nucleic acids -  
 Claim 27; Seq ID No 5932; 511pp; English.  
 The invention relates to antisense inhibitors of genes essential to  
 prokaryotic cellular proliferation, their use in identifying the  
 genes, their use in the discovery of novel antibiotics, the essential  
 genes themselves and the encoded proteins. The prokaryotes used are  
 Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella  
 pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The  
 invention is also useful for the identification of potential new targets  
 for antibiotic development. The antisense nucleic acids can also be used  
 to identify proteins used in proliferation, to express these proteins,  
 and to obtain antibodies capable of binding to the expressed proteins.  
 The proteins can be used to screen compounds in rational drug discovery  
 programmes. The antisense nucleic acid sequence is also useful to screen  
 for homologous nucleic acids which are required for cell proliferation in  
 a wide variety of organisms. The present sequence encodes an  
 essential prokaryotic cellular proliferation protein.  
 Note: The sequence data for this patent did not form part  
 of the printed specification, but was obtained in electronic  
 format directly from WIPO at

CC ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 1317 BP; 294 A; 323 C; 400 G; 300 T; 0 other;  
 Alignment Scores:  
 Pred. No.: 0.00126 Length: 1317  
 Score: 14.00 Matches: 14  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 3.12% Indels: 0  
 DB: 23 Gaps: 0

US-09-701-229-2 (1-448) x AAS52295 (1-1317)

QY 157 GluLeuTyRValLeuGluLeuSerSerPheGlnLeuGluThr 170  
 |||||  
 Db 457 GAACTGTACGTGCTGGAACTGTCGAGCTTCCAGCTGGAAACC 498

RESULT 10  
 AAS56341  
 ID AAS56341 standard; DNA; 1317 BP.  
 XX  
 AC AAS56341;  
 XX  
 DT 13-FEB-2002 (first entry)  
 XX  
 DE Salmonella typhi DNA for cellular proliferation protein #374.  
 XX  
 KW Antisense; ds; prokaryotic cellular proliferation gene;  
 KW antibiotic; antibacterial; drug design.  
 XX  
 OS Salmonella typhi.  
 XX  
 PN WO200170955-A2.  
 XX  
 PD 27-SEP-2001.  
 XX  
 PF 21-MAR-2001; 2001WO-US09180.  
 XX  
 PR 21-MAR-2000; 2000US-191078P.  
 PR 23-MAY-2000; 2000US-206848P.  
 PR 26-MAY-2000; 2000US-207727P.  
 PR 23-OCT-2000; 2000US-242578P.  
 PR 27-NOV-2000; 2000US-253625P.  
 PR 22-DEC-2000; 2000US-257931P.  
 PR 16-FEB-2001; 2001US-269308P.  
 XX  
 PA (ELIT-) ELITRA PHARM INC.  
 XX  
 PI Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;  
 Yamamoto RT, Xu HH;  
 WPI: 2001-611495/70.  
 P-PSDB; AAU38482.  
 New polynucleotides for the identification and development of  
 antibiotics, comprise sequences of antisense nucleic acids -  
 Claim 27; Seq ID No 9978; 511pp; English.  
 The invention relates to antisense inhibitors of genes essential to  
 prokaryotic cellular proliferation, their use in identifying the  
 genes, their use in the discovery of novel antibiotics, the essential  
 genes themselves and the encoded proteins. The prokaryotes used are  
 Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella  
 pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The  
 invention is also useful for the identification of potential new targets  
 for antibiotic development. The antisense nucleic acids can also be used  
 to identify proteins used in proliferation, to express these proteins,  
 and to obtain antibodies capable of binding to the expressed proteins.  
 The proteins can be used to screen compounds in rational drug discovery  
 programmes. The antisense nucleic acid sequence is also useful to screen  
 for homologous nucleic acids which are required for cell proliferation in

CC a wide variety of organisms. The present sequence encodes an  
CC essential prokaryotic cellular proliferation protein.  
CC Note: The sequence data for this patent did not form part  
CC of the printed specification, but was obtained in electronic  
CC format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences.

XX SQ Sequence 1317 BP; 258 A; 341 C; 427 G; 291 T; 0 other;

Alignment Scores:  
Pred. No.: 0.00126 Length: 1317  
Score: 14.00 Matches: 14  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 3.12% Indels: 0  
DB: 23 Gaps: 0

US-09-701-229-2 (1-448) x AAS56341 (1-1317)

QY 157 GluLeuTyrValLeuGluLeuSerPheGlnLeuGluThr 170  
AAS73012  
Db 457 GAATTGTACGTCTGGAATATATCCAGTTTCCCAATTGGAACT 498

RESULT 11

AAS73012  
ID AAS73012 standard; cDNA; 2058 BP.

XX AAS73012;

XX 13-FEB-2002 (first entry)

DE DNA encoding novel human diagnostic protein #8816.

KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.

OS Homo sapiens.

XX WO200175067-A2.

XX 11-OCT-2001.

XX 30-MAR-2001; 2001WO-US08631.

XX 31-MAR-2000; 2000US-0540217.

XX 23-AUG-2000; 2000US-0649167.

XX (HYSE-) HYSEQ INC.

PA Drmanac RT, Liu C, Tang YT;

PI WPI: 2001-639362/73.

DR P-PSDB; ABG08825.

XX New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity

XX Claim 1; SEQ ID No 8816; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.

CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. AAS64197-AAS94564 represent novel human  
CC diagnostic coding sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX SQ Sequence 2058 BP; 496 A; 526 C; 574 G; 462 T; 0 other;

Alignment Scores:  
Pred. No.: 0.00192 Length: 2058  
Score: 14.00 Matches: 14  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 3.12% Indels: 0  
DB: 23 Gaps: 0

US-09-701-229-2 (1-448) x AAS73012 (1-2058)

QY 157 GluLeuTyrValLeuGluLeuSerPheGlnLeuGluThr 170  
Db 1783 GAACGTACGTCTGGAACCTGCGAGCTTCCAGCTGGAACC 1824

RESULT 12

AAS69095

ID AAS69095 standard; cDNA; 3117 BP.

XX AAS69095;

XX 13-FEB-2002 (first entry)

DE DNA encoding novel human diagnostic protein #4899.

KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.

OS Homo sapiens.

XX WO200175067-A2.

XX 11-OCT-2001.

XX 30-MAR-2001; 2001WO-US08631.

XX 31-MAR-2000; 2000US-0540217.

XX 23-AUG-2000; 2000US-0649167.

XX (HYSE-) HYSEQ INC.

PA Drmanac RT, Liu C, Tang YT;

PI WPI: 2001-639362/73.

DR P-PSDB; ABG04908.

XX New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity

XX Claim 1; SEQ ID No 4899; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC quantitating a polypeptide in tissue, as molecular weight markers and as

CC a food supplement. (II) and its binding partners are useful in medical  
 CC imaging of sites expressing (II). (I) and (II) are useful for treating  
 CC disorders involving aberrant protein expression or biological activity.  
 CC The polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. AAS64197-AAS94564 represent novel human  
 CC diagnostic coding sequences of the invention.  
 CC Note: The sequence data for this patent did not appear in the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX SQ Sequence 3117 BP; 749 A; 820 C; 856 G; 692 T; 0 other;

Alignment Scores:  
 Pred. No.: 0.00284 Length: 3117  
 Score: 14.00 Matches: 14  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 3.12% Indels: 0  
 DB: 23 Gaps: 0

US-09-701-229-2 (1-448) x AAS69095 (1-3117)

OY 157 GluLeuTyrValLeuGluLeuSerSerPheGlnLeuGluThr 170  
 |||||  
 DB 355 GAACGTGACGTGCTGGAACGTGCGAGCTCCAGCTGGAACC 396

RESULT 13

AAS89240  
 ID AAS89240 standard; cDNA; 3117 BP.

XX AC AAS89240;

XX DT 13-FEB-2002 (first entry)

XX DE DNA encoding novel human diagnostic protein #25044.

XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
 XX food supplement; medical imaging; diagnostic; genetic disorder; ss.

XX OS Homo sapiens.

XX PN WO200175067-A2.

XX PD 11-OCT-2001.

XX PF 30-MAR-2001; 2001WO-US08631.

XX PR 31-MAR-2000; 2000US-0540217.

XX PR 23-AUG-2000; 2000US-0649167.

XX PA (HYSE-) HYSEQ INC.

XX PI Drmanac RT, Liu C, Tang YT;

XX DR WPI; 2001-639362/73.

XX DR P-PSDB; ABG25053.

XX PT New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity

XX PS Claim 1; SEQ ID No 25044; 103pp; English.

XX CC The invention relates to isolated polynucleotide (I) and  
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
 CC and gene mapping, and in recombinant production of (II). The  
 CC polynucleotides are also used in diagnostics as expressed sequence tags  
 CC for identifying expressed genes. (I) is useful in gene therapy techniques

CC to restore normal activity of (II) or to treat disease states involving  
 CC (II). (II) is useful for generating antibodies against it, detecting or  
 CC quantitating a polypeptide in tissue, as molecular weight markers and as  
 CC a food supplement. (II) and its binding partners are useful in medical  
 CC imaging of sites expressing (II). (I) and (II) are useful for treating  
 CC disorders involving aberrant protein expression or biological activity.  
 CC The polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. AAS64197-AAS94564 represent novel human  
 CC diagnostic coding sequences of the invention.  
 CC Note: The sequence data for this patent did not appear in the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX SQ Sequence 3117 BP; 749 A; 820 C; 856 G; 692 T; 0 other;

Alignment Scores:  
 Pred. No.: 0.00284 Length: 3117  
 Score: 14.00 Matches: 14  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 3.12% Indels: 0  
 DB: 23 Gaps: 0

US-09-701-229-2 (1-448) x AAS89240 (1-3117)

OY 157 GluLeuTyrValLeuGluLeuSerSerPheGlnLeuGluThr 170  
 |||||  
 DB 355 GAACGTGACGTGCTGGAACGTGCGAGCTCCAGCTGGAACC 396

RESULT 14

AAS94542  
 ID AAS94542 standard; cDNA; 4831 BP.

XX AC AAS94542;

XX DT 13-FEB-2002 (first entry)

XX DE DNA encoding novel human diagnostic protein #30346.

XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
 XX food supplement; medical imaging; diagnostic; genetic disorder; ss.

XX OS Homo sapiens.

XX PN WO200175067-A2.

XX PD 11-OCT-2001.

XX PF 30-MAR-2001; 2001WO-US08631.

XX PR 31-MAR-2000; 2000US-0540217.

XX PR 23-AUG-2000; 2000US-0649167.

XX PA (HYSE-) HYSEQ INC.

XX PI Drmanac RT, Liu C, Tang YT;

XX DR WPI; 2001-639362/73.

XX DR P-PSDB; ABG30355.

XX PT New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity

XX PS Claim 1; SEQ ID No 30346; 103pp; English.

XX CC The invention relates to isolated polynucleotide (I) and  
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
 CC and gene mapping, and in recombinant production of (II). The  
 CC polynucleotides are also used in diagnostics as expressed sequence tags  
 CC for identifying expressed genes. (I) is useful in gene therapy techniques

CC and gene mapping, and in recombinant production of (II). The  
 CC polynucleotides are also used in diagnostics as expressed sequence tags  
 CC for identifying expressed genes. (I) is useful in gene therapy techniques  
 CC to restore normal activity of (II) or to treat disease states involving  
 CC (II). (II) is useful for generating antibodies against it, detecting or  
 CC quantitating a polypeptide in tissue, as molecular weight markers and as  
 CC a food supplement. (II) and its binding partners are useful in medical  
 CC imaging of sites expressing (II). (I) and (II) are useful for treating  
 CC disorders involving aberrant protein expression or biological activity.  
 CC The polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. AAS64197-AAS94564 represent novel human  
 CC diagnostic coding sequences of the invention.  
 CC Note: The sequence data for this patent did not appear in the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX SQ Sequence 4831 BP; 1180 A; 1249 C; 1353 G; 1049 T; 0 other;

#### Alignment Scores:

Pred. No.:	0.00429	Length:	4831
Score:	14.00	Matches:	14
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	3.12%	Indels:	0
DB:	23	Gaps:	0

US-09-701-229-2 (1-448) x AAS94564 (1-4831)

QY 157 GluLeuTyrrValLeuGluLeuSerSerPheGlnLeuGluThr 170

DB 4556 GAACGTACGTGCTGGAACCTGCGAGCTTCAGCTGGAACCC 4597

#### RESULT 15

AAS89816

ID AAS89816 standard; cDNA; 4942 BP.

XX AC AAS89816;

XX DT 13-FEB-2002 (first entry)

XX DE DNA encoding novel human diagnostic protein #25620.

XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
 KW food supplement; medical imaging; diagnostic; genetic disorder; ss.

XX OS Homo sapiens.

XX PN WO200175067-A2.

XX PD 11-OCT-2001.

XX PF 30-MAR-2001; 2001WO-US08631.

XX PR 31-MAR-2000; 2000US-0540217.

XX PR 23-AUG-2000; 2000US-0649167.

XX PA (HYSE-) HYSEQ INC.

XX PI Drmanac RT, Liu C, Tang YT;

XX DR WPI; 2001-639362/73.

XX DR P-PSDB; ABG25629.

XX PT New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity

XX PS Claim 1; SEQ ID No 25620; 103pp; English.

XX

CC The invention relates to isolated polynucleotide (I) and  
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
 CC and gene mapping, and in recombinant production of (II). The  
 CC polynucleotides are also used in diagnostics as expressed sequence tags  
 CC for identifying expressed genes. (I) is useful in gene therapy techniques  
 CC to restore normal activity of (II) or to treat disease states involving  
 CC (II). (II) is useful for generating antibodies against it, detecting or  
 CC quantitating a polypeptide in tissue, as molecular weight markers and as  
 CC a food supplement. (II) and its binding partners are useful in medical  
 CC imaging of sites expressing (II). (I) and (II) are useful for treating  
 CC disorders involving aberrant protein expression or biological activity.  
 CC The polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. AAS64197-AAS94564 represent novel human  
 CC diagnostic coding sequences of the invention.  
 CC Note: The sequence data for this patent did not appear in the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX SQ Sequence 4942 BP; 1109 A; 1296 C; 1387 G; 1150 T; 0 other;

#### Alignment Scores:

Pred. No.:	0.00439	Length:	4942
Score:	14.00	Matches:	14
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	3.12%	Indels:	0
DB:	23	Gaps:	0

US-09-701-229-2 (1-448) x AAS89816 (1-4942)

QY 157 GluLeuTyrrValLeuGluLeuSerSerPheGlnLeuGluThr 170

DB 2179 GAACGTACGTGCTGGAACCTGCGAGCTTCAGCTGGAACCC 2220

Search completed: August 14, 2003, 09:24:59  
 Job time : 395 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus.p2n model

Run on: August 14, 2003, 07:13:51 ; Search time 349 seconds  
(without alignments)  
2872.772 Million cell updates/sec

Title: US-09-701-229-2

Perfect score: 2260

Sequence: 1 MSIIASHDFRIVVGLSGM.....MFKNFEEGRLEFAKAVEELA 448

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5

Ygapop 10.0 , Ygapext 0.5

Fgapop 6.0 , Fgapext 7.0

Delop 6.0 , Delext 7.0

Searched: 1504479 seqs, 1118970152 residues

Total number of hits satisfying chosen parameters: 2626304

Minimum DB seq length: 25

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame.p2n.model -DEV=xl

-Q=/cgn2\_1/USPTO.spool/US0701229/runat\_09082003\_161159\_4434/app\_query.fasta\_1.647

-DB=PublishedApplications\_NA -QFMT=fastap -SUFFIX=p2n.rnpb -MINMATCH=0.1

-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62

-TRANS=human40.cdi -LIST=45 -DOALIGN=200 -THR\_SCORE=pct -THR\_MAX=100

-THR\_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500

-MINLEN=25 -MAXLEN=2000000000

-USER=US0701229.qcgn.1.1.333 @runat\_09082003\_161159\_4434 -NCPU=6 -ICPU=3

-NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG

-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPEXT=10 -XGAPEXT=0.5 -FGAPOP=6

-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications\_NA:

- 1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq:\*
- 2: /cgn2\_6/ptodata/1/pubpna/PCT\_NEW\_PUB.seq:\*
- 3: /cgn2\_6/ptodata/1/pubpna/US06\_NEW\_PUB.seq:\*
- 4: /cgn2\_6/ptodata/1/pubpna/US06\_PUBCOMB.seq:\*
- 5: /cgn2\_6/ptodata/1/pubpna/US07\_NEW\_PUB.seq:\*
- 6: /cgn2\_6/ptodata/1/pubpna/PCTUS\_PUBCOMB.seq:\*
- 7: /cgn2\_6/ptodata/1/pubpna/US08\_NEW\_PUB.seq:\*
- 8: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq:\*
- 9: /cgn2\_6/ptodata/1/pubpna/US09A\_PUBCOMB.seq:\*
- 10: /cgn2\_6/ptodata/1/pubpna/US09B\_PUBCOMB.seq:\*
- 11: /cgn2\_6/ptodata/1/pubpna/US09C\_PUBCOMB.seq:\*
- 12: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq:\*
- 13: /cgn2\_6/ptodata/1/pubpna/US10A\_PUBCOMB.seq:\*
- 14: /cgn2\_6/ptodata/1/pubpna/US10B\_PUBCOMB.seq:\*
- 15: /cgn2\_6/ptodata/1/pubpna/US10\_NEW\_PUB.seq:\*
- 16: /cgn2\_6/ptodata/1/pubpna/US60\_NEW\_PUB.seq:\*
- 17: /cgn2\_6/ptodata/1/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Length	ID	Description
------------	-------------	--------	----	-------------

1	2260	100.0	1347	9	US-09-815-242-7912	Sequence 7912, Ap
2	989.5	43.8	1317	9	US-09-815-242-9978	Sequence 9978, Ap
3	985.5	43.6	1317	9	US-09-741-669-259	Sequence 259, App
4	985.5	43.6	1317	9	US-09-815-242-5932	Sequence 5932, Ap
5	879	38.9	1314	9	US-09-815-242-7090	Sequence 7090, Ap
6	879	38.9	1830121	14	US-10-329-960-1	Sequence 1, Appli
c	716	31.7	640681	10	US-09-790-988-1	Sequence 1, Appli
7	618.5	27.4	1401	14	US-10-156-761-6097	Sequence 1, Appli
8	618.5	27.4	9025608	14	US-10-156-761-6097	Sequence 9535, Ap
9	592.5	26.2	1353	9	US-09-815-242-9535	Sequence 9300, Ap
10	586.5	26.0	1353	9	US-09-815-242-9300	Sequence 1685, Ap
11	534.5	23.7	1352	10	US-09-974-300-1685	Sequence 79, Appl
12	525	23.2	1461	10	US-09-712-363-79	Sequence 3336, Ap
13	519	23.0	1368	9	US-09-815-242-3936	Sequence 6539, Ap
14	519	23.0	1371	9	US-09-815-242-6539	Sequence 109, App
15	493	21.8	1003	9	US-09-765-272-109	Sequence 4419, Ap
16	470	20.8	1347	9	US-09-815-242-4419	Sequence 8250, Ap
17	470	20.8	1350	9	US-09-815-242-8250	Sequence 8695, Ap
18	470	20.8	1350	9	US-09-815-242-8695	Sequence 1, Appli
19	463.5	20.5	3309400	10	US-09-738-626-1	Sequence 128, App
c	370.5	16.4	32768	10	US-09-070-927A-128	Sequence 7232, Ap
20	329	14.6	1269	9	US-09-815-242-7232	Sequence 80, Appl
21	324.5	14.4	1320	8	US-08-781-986A-80	Sequence 2193, Ap
22	259.5	11.5	1392	14	US-10-156-761-1	Sequence 1, Appli
23	259.5	11.5	9025608	14	US-10-156-761-1	Sequence 7914, Ap
c	232.5	10.3	1377	9	US-09-815-242-7914	Sequence 257, App
24	232.5	9.8	1359	9	US-09-741-669-257	Sequence 257, App
25	220.5	9.8	1359	9	US-09-815-242-5930	Sequence 5930, Ap
26	220.5	9.8	1359	9	US-09-712-363-82	Sequence 82, Appl
27	213	9.4	1608	10	US-09-738-626-2370	Sequence 2365, Ap
28	203	9.0	468	10	US-09-738-626-2365	Sequence 9976, Ap
29	199.5	8.8	1359	9	US-09-815-242-9976	Sequence 7093, Ap
30	197	8.7	1428	9	US-09-815-242-7093	Sequence 7910, Ap
31	195.5	8.7	1443	9	US-09-815-242-7910	Sequence 9975, Ap
32	192	8.5	1455	9	US-09-815-242-9975	Sequence 2370, Ap
33	192	8.5	1551	10	US-09-738-626-2370	Sequence 2362, Ap
34	183	8.1	1458	10	US-09-738-626-2362	Sequence 76, Appl
35	183	8.1	1485	10	US-09-712-363-76	Sequence 2886, Ap
36	182.5	8.1	717	9	US-09-939-980-78	Sequence 6095, Ap
37	179	7.9	383	9	US-09-815-242-2986	Sequence 262, App
38	177.5	7.9	1413	14	US-10-156-761-6095	Sequence 5935, Ap
c	175	7.7	1476	9	US-09-741-669-262	Sequence 9981, Ap
39	175	7.7	1476	9	US-09-815-242-5935	Sequence 3822, Ap
40	174.5	7.7	1476	9	US-09-815-242-9981	Sequence 6676, Ap
41	174	7.7	1362	9	US-09-815-242-3822	
42	174	7.7	1425	9	US-09-815-242-6676	
43	174	7.7	1425	9	US-09-815-242-6676	
44	174	7.7	1425	9	US-09-815-242-6676	
45	174	7.7	1425	9	US-09-815-242-6676	

ALIGNMENTS

RESULT 1

US-09-815-242-7912  
; Sequence 7912, Application US/09815242  
; Patent No. US20020061569A1  
; GENERAL INFORMATION:  
; APPLICANT: Hasebeck, Robert  
; APPLICANT: Ohlsen, Karl L.  
; APPLICANT: Zyskind, Judith W.  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John D.  
; APPLICANT: Carr, Grant J.  
; APPLICANT: Yamamoto, Robert T.  
; APPLICANT: Xu, H. Howard  
; TITLE OF INVENTION: Identification of Essential Genes in  
; FILE OF INVENTION: Prokaryotes  
; FILE REFERENCE: ELITRA.011A  
; CURRENT APPLICATION NUMBER: US/09/815.242  
; CURRENT FILING DATE: 2001-03-21  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26

```
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7912
; LENGTH: 1347
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1347)
US-09-815-242-7912

Alignment Scores:
Pred. No.: 2,29e-258 Length: 1347
Score: 2260.00 Matches: 448
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 9 Gaps: 0

US-09-701-229-2 (1-448) x US-09-815-242-7912 (1-1347)

QY 1 MetSerLeuLeuAlaSerAspHisPheArgIleValValGlyLeuGlyLysSerGlyMet 20
DB 1 ATGAGCCTGATCGCTCCGACCACTCCGCATCGTTGCGGCTCGGCAAGAGCGGCATG 60

QY 21 SerLeuValArgTyrLeuAlaArgArgGlyLeuProPheAlaValValAspThrArgGlu 40
DB 61 TCCCTGGTGCCTACTGCGCGCGCGGCTTGCCTTCGCGGTGTCGATACCGAGAG 120

QY 41 AsnProProGluLeuAlaThrLeuArgAlaGlnTyrProGlnValGluValArgCysGly 60
DB 121 AACCCGCGGAGCTGGCCACCTGCTGCGCCAGTATCGCAGGTGGAAAGTGGCTGCGGC 180

QY 61 GluLeuAspAlaGluPheLeuCysSerAlaArgGluLeuTyrValSerProGlyLeuSer 80
DB 181 GAACTCGAGCGCGAGTCTCTGCTCCGCCCGCAACTATGTCAGCCCGCGCTGTGCG 240

QY 81 LeuArgThrProAlaLeuValGlnAlaAlaLysGlyValArgIleSerGlyAspIle 100
DB 241 CTGCGCACCCCTGCTGCTACAGCGCGCGGAAAGCGTGCATCTCCGGTGACATC 300

QY 101 AspLeuPheAlaArgGluAlaLysAlaProIleValAlaIleThrGlySerAsnAlaLys 120
DB 301 GATCTCTTGGCCCGGAGCGGAAGCGCCGATCGTCGCATCATCCCGGTTCACCGCAG 360

QY 121 SerThrValThrLeuValGlyGluMetAlaValAlaAlaAspLysArgValAlaVal 140
DB 361 AGCACCTGACACCTGTTGGCGCAATGCGGTGGCGGCAACGGTGTGCCCGTC 420

QY 141 GlyGlyAsnLeuGlyThrProAlaLeuAspLeuAlaAspAspIleGluLeuTyrVal 160
DB 421 GCGCGCAACCTCGGCACCCCGCGCTCGACCTGCTGGCGGACGACATCGAGCTGTAC 480

QY 161 LeuGluLeuSerSerPheGlnLeuGluThrCysAspArgLeuAsnAlaGluValAlaThr 180
DB 481 TTGGAGCTGTGAGCTTCCAGTGGAAACCTGCGATCGCTCAACCGCGAGGTGCGGACC 540

QY 181 ValLeuAsnValSerGluAspHisMetAspArgTyrAspGlyMetAlaAspTyrHisLeu 200
DB 541 GTGCTGACCTGACGAGACCATATGATGCTACGACGGCATGGCTGACTACCACTG 600

QY 201 AlaLysHisArgIlePheArgGlyAlaArgGlnValValAlaAsnArgAlaAspAlaLeu 220
DB 601 GCCAAGCACCGGATCTTCCGCGGTGCCCGCAGGTGCTGTGAATCGCGCGCATGCCCTG 660
```

```
QY 221 ThrArgProLeuLeuAlaAspThrValProCysTyrTrpSerPheGlyLeuAsnLysProAsp 240
DB 661 ACCGACCGCTGATCGCGCATACCGTGCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 720

QY 241 PheLysAlaPheGlyLeuIleGluGluAspGlyGlnLysTrpLeuAlaPheGlnPheAsp 260
DB 721 TTCAAGGCTTTCGGCTGATCGAGGAAGACGCCAGAGTGGCTGGCTGCTGCTGCTGCTGCTG 780

QY 261 LysLeuLeuProValGlyGluLeuLysIleArgGlyAlaHisAsnTyrSerAsnAlaLeu 280
DB 781 AAGCTGCTGCGGTTGGGAACCTGAAGATCCGTGGCGCCACCACTATTTCACACGCGCTC 840

QY 281 AlaAlaLeuAlaLeuGlyHisAlaValGlyLeuProPheAspAlaMetLeuGlyAlaLeu 300
DB 841 GCGCGCTGCGCTGGCGCATCGGTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTG 900

QY 301 LysAlaPheSerGlyLeuAlaHisArgCysGlnTyrValArgGluArgGlnGlyValSer 320
DB 901 AAGGCTTTCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 960

QY 321 TyrTyrAspSerLysAlaThrAsnValGlyAlaAlaLeuAlaAlaIleGluGlyLeu 340
DB 961 TACTACGACGATTCACAGGCCACCAACGTCGCGCGCGCTGCGCGCGATCGAGGGCTG 1020

QY 341 GlyAlaAspIleAspGlyLysLeuValLeuAlaGlyLysAspGlyLysGlyAlaAsp 360
DB 1021 GGTCCGACATCGACGCGCAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1080

QY 361 PheHisAspLeuArgGluProValAlaArgPheCysArgAlaValValLeuLeuGlyArg 380
DB 1081 TTCATGACCTGCGGCGAGCGGTCGCGGCTTCTGCGGCGCGGTGGTACTGCTGGCGCT 1140

QY 381 AspAlaGlyLeuLeuAlaGlnAlaLeuGlyAsnAlaValProLeuValArgValAlaThr 400
DB 1141 GACGCGGCTGATTCGCCGACCTGGGCAACGCGGTACCGCTGGTGGCTGCGCAACG 1200

QY 401 LeuAspGluAlaValArgGlnAlaAlaGluLeuAlaArgGluGlyAspAlaValLeuLeu 420
DB 1201 CTGGACGAAGCAGTCCGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1260

QY 421 SerProAlaCysAlaSerLeuAspMetPheLysAsnPheGluGluArgGlyArgLeuPhe 440
DB 1261 TCGCGCGCTGCGCGAGCGCTGGACATGTTCAAGAACTTCGAAGAACGCGGACGCTGTTC 1320

QY 441 AlaLysAlaValGluLeuAla 448
DB 1321 GCCAAGCGTAGAGGAGTAGCG 1344
```

## RESULT 2

```
US-09-815-242-9978
; Sequence 9978, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
```

; PRIOR FILING DATE: 2000-10-23  
 ; PRIOR APPLICATION NUMBER: 60/253,625  
 ; PRIOR FILING DATE: 2000-11-27  
 ; PRIOR APPLICATION NUMBER: 60/257,931  
 ; PRIOR FILING DATE: 2000-12-22  
 ; PRIOR APPLICATION NUMBER: 60/269,308  
 ; PRIOR FILING DATE: 2001-02-16  
 ; NUMBER OF SEQ ID NOS: 14110  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 9978  
 ; LENGTH: 1317  
 ; TYPE: DNA  
 ; ORGANISM: Salmonella typhi  
 ; FEATURE:  
 ; NAME/KEY: CDS  
 ; LOCATION: (1)...(1317)  
 US-09-815-242-9978

Alignment Scores:  
 Pred. No.: 1,7e-107 Length: 1317  
 Score: 989.50 Matches: 230  
 Percent Similarity: 64.56% Conservative: 56  
 Best Local Similarity: 51.92% Mismatches: 136  
 Query Match: 43.78% Indels: 21  
 DB: 9 Gaps: 9

US-09-701-229-2 (1-448) x US-09-815-242-9978 (1-1317)

QY 11 IleValValGlyLeuGlyLysSerGlyMetSerLeuValArgTyrLeuAlaArgGly 30  
 DB 28 GTCAATATATCGCTGGCGCTTAACCGGACTCTCTTCGCTGGACTTTTCTCGCGCGCGC 87  
 QY 31 LeuProPheLaValValAspThrArgGluAsnProProGluLeuAlaThrLeuArgAla 50  
 DB 88 GTGACGCGCGGGGTGGTACTCGCTGACGCGCGCGGTCTGTGTAAGCTG----- 141  
 QY 51 GlnTyrProGln-----ValGluValArgCysGlyGluLeuAspAlaGluPheLeuCysSer 69  
 DB 142 -----CGCAAGAGGTGAGCGTCACGTTGGCGCGCTGAACGACGAGTGGCTTTAGCG 195  
 QY 70 AlaArgGluLeuTyrValSerProGlyLeuSerLeuArgThrProAlaLeuValGluAla 89  
 DB 196 GCGGATTTAATCGTCGCCAGCCCTGGTATTCCTGGCGCATCCCTCTCTAGCGCTGCC 255  
 QY 90 AlaAlaLysGlyValArgIleSerGlyAspIleAspLeuPheAlaArgGluAlaLysAla 109  
 DB 256 GCCAGCGCTGGAGTGAATTTGCGCGGATATCGAACTGTTTTCGCGGAGCAAGCG 315  
 QY 110 ProIleValAlaIleThrGlySerAsnAlaLysSerThrValThrLeuValGlyGlu 129  
 DB 316 CCGATTGTGGCCATCAGCGGCTCGAAGCGCAAAAGCACCGTACACCTTATGTGGCGGAG 375  
 QY 130 MetAlaValAlaAspLysArgValAlaValGlyGlyAsnLeuGlyThrProAlaLeu 149  
 DB 376 ATGGCAAGCGCGCGCGCTCAATGTGGCGTTCGGCGCAATATCGTCTTCGCGCGCTG 435  
 QY 150 AspLeuLeuAlaAspAspIleGluLeuTyrValLeuGluLeuSerSerPheGlnLeuGlu 169  
 DB 436 ATGCTGCTGGATCGCATCGCAATTTGACGTGTGGAATTTATTCAGTTTCCATTTGGA 495  
 QY 170 ThrCysAspArgLeuAsnAlaGluValAlaThrValLeuAsnValSerGluAspHisMet 189  
 DB 496 ACTACCTCAAGTTTTCGAAGCGCGCGCGCAACGCTGCTCAACGCTCAAGATCATATG 555  
 QY 190 AspArgTyrAsp---GlyMetAlaAspTyrHisLeuAlaLysHisArgIlePheArgGly 208  
 DB 556 GACCGGTATCCGTTGGTTTGCACAGTACCGCGCGGCAAACTCGCGCTCTACAGAAG 615  
 QY 209 AlaArgGlnValValAlaAsnArgAlaAspAlaLeuThrArgProLeuIleAlaAspThr 228  
 DB 616 GCGAAAGTGCCTAGTAGTGAATGCCGATGACGCGTTGACGATGCGCGGTACGCGGCGCAT 675  
 QY 229 ValProCysTyrPheGlyLeuAsnLysProAspPheLysAlaPheGlyLeuIleGlu 248

DB 676 GAGCGCTGCGTCAGCTTTGGCGTCAATATGGTGGTATATAC-----CTTAATCGT 726  
 QY 249 GluAspGlyGlnLysTrpLeuAlaPheGlnPheAspLysLeuLeuProValGlyGluLeu 268  
 DB 727 CAGCAGGCGCAACCTGGCTACGAGTCAAGAGTGAAGAGTGGCTGAACAGAGAGATG 786  
 QY 269 LysIleArgGlyAlaHisAsnTyrSerAsnAlaLeuAlaLeuAlaLeuGlyHisAla 288  
 DB 787 AAGCTTCCGCTGACATATACCAATCGTTCGCGCGCTGGCGCTGGCGGATGCC 846  
 QY 289 ValGlyLeuProPheAspAlaMetLeuGlyAlaLeuLysAlaPheSerGlyLeuAlaHis 308  
 DB 847 GTAGTCTGCGCGCGCGCAGCAGTTTGAAGCGGTTCAGCAGATTTACCGGTCTGGCGCAT 906  
 QY 309 ArgCysGlnTrpValArgGluArgGlnGlyValSerTyrTyrAspAspSerLysAlaThr 328  
 DB 907 CGCTTCCAGCTGGCGTGGAGCATACCGCGCTTCGTTGGATCAACAGCTCGAAAGCGACC 966  
 QY 329 AsnValGlyAlaAlaLeuAlaAlaIleGluGlyLeuGlyAlaAspIleAspGlyLysLeu 348  
 DB 967 AATGTCGCGCAGTACCGAGCGCGCTAAACGGTTTG-----CATGTGACGGTACGCTG 1020  
 QY 349 ValLeuLeuAlaGlyGlyAspGlyLysGlyAlaAspPheHisAspLeuArgGluProVal 368  
 DB 1021 CATCTGCTGCTGGCGCGCGCAGCTAAGTTCGCGCAGACTTT-----TCTCCGCTG 1068  
 QY 369 AlaArgPheCysArgAla-----ValValLeuLeuGlyArgAspAlaGlyLeu 384  
 DB 1069 GCGCGCATATCGACCGCGGATCGTATCCGCTGATTCGTTGGCGCGATGGCGCGCAG 1128  
 QY 385 IleAlaGlnAlaLeuGlyAsnAlaValProLeuValArgValAlaThrLeuAspGluAla 404  
 DB 1129 CTTCGCC---GCATGCGCTCGCGAAATC---GCCCAACAGACTGAGACGATGAAGAGCG 1182  
 QY 405 ValArgGlnAlaAlaGluLeuAlaArgGluGlyAspAlaValLeuLeuSerProHisCys 424  
 DB 1183 ATGCGTTTGTGGCGCGCGCACGTTTCAGCGCGTGTATGTTGCTGCTTCGCCCGCTGC 1242  
 QY 425 AlaSerLeuAspMetPheLysAsnPheGluGluArgGlyArgLeuPheAlaLysAlaVal 444  
 DB 1243 GCGACGCTCATGACTTTAAANATTTGAGCAACGGCGGATGCTTTACCGCTGCGC 1302  
 QY 445 GluGluLeu 447  
 DB 1303 AAGAGTTA 1311

# RESULT 3

; US-09-741-669-259  
 ; Sequence 259, Application US/09741669  
 ; Patent No. US20020022718A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Forsyth, R. Allyn  
 ; APPLICANT: Ohlsen, Kari L.  
 ; APPLICANT: Zyskind, Judith W.  
 ; TITLE OF INVENTION: Genes identified as required for  
 ; proliferation of E. coli  
 ; FILE REFERENCE: ELITRA.009A  
 ; CURRENT APPLICATION NUMBER: US/09/741,669  
 ; CURRENT FILING DATE: 2000-12-19  
 ; PRIOR APPLICATION NUMBER: US 60/173005  
 ; PRIOR FILING DATE: 1999-12-23  
 ; NUMBER OF SEQ ID NOS: 481  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 259  
 ; LENGTH: 1317  
 ; TYPE: DNA  
 ; ORGANISM: Escherichia coli  
 ; FEATURE:  
 ; NAME/KEY: CDS  
 ; LOCATION: (1)...(1317)  
 US-09-741-669-259

Alignment Scores:

```

Pred. No.: 5.09e-107 Length: 1317
Score: 985.50 Matches: 227
Percent Similarity: 64.33% Conservative: 58
Best Local Similarity: 51.24% Mismatches: 137
Query Match: 43.61% Indels: 21
DB: 9 Gaps: 9

US-09-701-229-2 (1-448) x US-09-741-669-259 (1-1317)

QY 11 IleValValGlyLeuGlySerGlyMetSerLeuValArgTyrLeuAlaArgGly 30
Db : ::::::::::::::::::::: ||| ::|
QY 28 GTCATTATCGGCTGGCCCTCCAGCGGCTTCCTCGGTGAGCTTTTCCTCGCTCGCGGT 87
Db : ::::::::::::::::::::: ||| ::|
QY 31 LeuProPheAlaValValAspThrArgGluAsnProProGluLeuAlaThrLeuArgAla 50
Db : ::::::::::::::::::::: ||| ::|
QY 88 GTGACGCCCGGCTTATGATACCGGTATGACACCGCTGGCTGGCTGATTAATA----- 141
Db : ::::::::::::::::::::: ||| ::|
QY 51 GlnTyrProGln---ValGluValArgCysGlyLeuAlaGluPheLeuCysSer 69
Db : ::::::::::::::::::::: ||| ::|
QY 142 -----CCCGAAGCCGTAGAACGCCACACCGCGAGTCTGAATGATGATGGCTGATGGCG 195
Db : ::::::::::::::::::::: ||| ::|
QY 70 AlaArgGluLeuTyrValSerProGlyLeuSerLeuArgThrProAlaLeuValGlnAla 89
Db : ::::::::::::::::::::: ||| ::|
QY 196 GCAGATCTGATGTGCGCCAGTCCCGGTATTGACATGGCGCATCCATCTTAAGCGCTGCC 255
Db : ::::::::::::::::::::: ||| ::|
QY 90 AlaAlaLysGlyValArgIleSerGlyAspIleAspLeuPheAlaArgGluAlaLysAla 109
Db : ::::::::::::::::::::: ||| ::|
QY 256 GCTGATGCCGGAATCGTTCGGGATATCGAGCTGTTCTCTCGCGAACAACGCA 315
Db : ::::::::::::::::::::: ||| ::|
QY 110 ProIleValAlaIleThrGlySerAsnAlaLysSerThrValThrThrLeuValGlu 129
Db : ::::::::::::::::::::: ||| ::|
QY 316 CGGATTGTGGCATTACCGGTTCTAACGCCAAAGCAGCGTCCACCGCTAGTGGTGAA 375
Db : ::::::::::::::::::::: ||| ::|
QY 130 MetAlaValAlaAlaAspLysArgValAlaValGlyGlyAsnLeuGlyThrProAlaLeu 149
Db : ::::::::::::::::::::: ||| ::|
QY 376 ATGGCGAAGCCGCGGGGTTAACGTGTGGTGGGCAATATTGGCGTGGCTGGCTG 435
Db : ::::::::::::::::::::: ||| ::|
QY 150 AspLeuLeuAlaAspIleGluLeuTyrValLeuGluLeuSerSerPheGlnLeuGlu 169
Db : ::::::::::::::::::::: ||| ::|
QY 436 ATGCTACTGGATGATGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 495
Db : ::::::::::::::::::::: ||| ::|
QY 170 ThrCysAspArgLeuAsnAlaGluValAlaThrValLeuAsnValSerGluAspHisMet 189
Db : ::::::::::::::::::::: ||| ::|
QY 496 ACCACCTCCAGCTTACAGCGGTAGCAGCGACCATCTGAAGCTGACTGAAGTCAATG 555
Db : ::::::::::::::::::::: ||| ::|
QY 190 AspArgTyrAsp---GlyMetAlaAspTyrHisLeuAlaLysHisArgIlePheArgGly 208
Db : ::::::::::::::::::::: ||| ::|
QY 556 GATCGCTATCGGTTGGTTTACACAGTATCGTCAGCAAAACTCGCATTTTACGAAAC 615
Db : ::::::::::::::::::::: ||| ::|
QY 209 AlaArgGlnValValAsnArgAlaAspAlaLeuThrArgProLeuIleAlaAspThr 228
Db : ::::::::::::::::::::: ||| ::|
QY 616 GCGAAGTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 675
Db : ::::::::::::::::::::: ||| ::|
QY 229 ValProCysTyrSerPheGlyLeuAsnLysProAspPheLysAlaPheClyLeuIleGlu 248
Db : ::::::::::::::::::::: ||| ::|
QY 676 GAACGCTCGCTGAGCTTGGCGTCAACATGGGTGACTATCAC-----CTGAATCAT 726
Db : ::::::::::::::::::::: ||| ::|
QY 249 GluAspGlyGlnLysTyrPheAlaPheGlnPheAspLysLeuLeuProValGlyGluLeu 268
Db : ::::::::::::::::::::: ||| ::|
QY 727 CAGCAGCGGGAACCTGGCTGGGTTAAAGCGGAGAAAGTCTGAATGTGAAGAGATG 786
Db : ::::::::::::::::::::: ||| ::|
QY 269 LysIleArgGlyAlaHisAsnTyrSerAsnAlaLeuAlaLeuAlaLeuGlyHisAla 288
Db : ::::::::::::::::::::: ||| ::|
QY 787 AAACCTTCGGGCAGCAATAACTACACCAATGGCTGGCGGCGCTGGCGCTGCAGATGCT 846
Db : ::::::::::::::::::::: ||| ::|
QY 289 ValGlyLeuProPheAspAlaMetLeuGlyAlaLeuLysAlaPheSerGlyLeuAlaHis 308
Db : ::::::::::::::::::::: ||| ::|
QY 847 GCAGGGTTACCGCGTGGCGGAGCGCTGAACCGGTTAACCCACATTCACGTGGTCTGCCGAT 906
Db : ::::::::::::::::::::: ||| ::|
QY 309 ArgCysGlnTyrValArgGlnGlyValSerTyrTyrAspSerLysAlaThr 328
Db : ::::::::::::::::::::: ||| ::|
QY 907 CGCTTTGAAGTTGTCTGGACGATACCGCGTACCTGCTGGATTAACGATTCAAGACCGACC 966

```

RESULT 4

```

US-09-815-242-5932
; Sequence 5932, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5932
; LENGTH: 1317
; TYPE: DNA
; ORGANISM: Escherichia coli
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1317)
US-09-815-242-5932
Alignment Scores:

```

1314



OTHER INFORMATION: n equals a, t, g or c  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (36636)..(36636)  
OTHER INFORMATION: n equals a, t, g or c  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (40808)..(40810)  
OTHER INFORMATION: n equals a, t, g or c  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (44416)..(44416)  
OTHER INFORMATION: n equals a, t, g or c  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (44905)..(44905)  
OTHER INFORMATION: n equals a, t, g or c  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (44975)..(44975)  
OTHER INFORMATION: n equals a, t, g or c  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (45593)..(45593)  
OTHER INFORMATION: n equals a, t, g or c  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (45732)..(45732)  
OTHER INFORMATION: n equals a, t, g or c  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (47036)..(47036)  
OTHER INFORMATION: n equals a, t, g or c  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (51334)..(51334)  
OTHER INFORMATION: n equals a, t, g or c  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (51602)..(51602)  
OTHER INFORMATION: n equals a, t, g or c  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (51786)..(51786)  
OTHER INFORMATION: n equals a, t, g or c  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (51805)..(51805)  
OTHER INFORMATION: n equals a, t, g or c  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (55369)..(55369)  
OTHER INFORMATION: n equals a, t, g or c  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (65309)..(65309)  
OTHER INFORMATION: n equals a, t, g or c  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (65313)..(65313)  
OTHER INFORMATION: n equals a, t, g or c  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (80024)..(80024)  
OTHER INFORMATION: n equals a, t, g or c  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (100091)..(100091)  
OTHER INFORMATION: n equals a, t, g or c  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (102696)..(102696)  
OTHER INFORMATION: n equals a, t, g or c

FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (105121)..(105121)  
OTHER INFORMATION: n equals a, t, g or c  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (107248)..(107248)  
OTHER INFORMATION: n equals a, t, g or c  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (117136)..(117136)  
OTHER INFORMATION: n equals a, t, g or c  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (119750)..(119750)  
OTHER INFORMATION: n equals a, t, g or c  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (119924)..(119924)  
OTHER INFORMATION: n equals a, t, g or c  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (120038)..(120038)  
OTHER INFORMATION: n equals a, t, g or c  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (121344)..(121344)  
OTHER INFORMATION: n equals a, t, g or c  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (122167)..(122167)  
OTHER INFORMATION: n equals a, t, g or c  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (122336)..(122336)  
OTHER INFORMATION: n equals a, t, g or c  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (131340)..(131340)  
OTHER INFORMATION: n equals a, t, g or c  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (131360)..(131360)  
OTHER INFORMATION: n equals a, t, g or c  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (139910)..(139910)  
OTHER INFORMATION: n equals a, t, g or c  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (140398)..(140398)  
OTHER INFORMATION: n equals a, t, g or c  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (142750)..(142750)  
OTHER INFORMATION: n equals a, t, g or c  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (145058)..(145058)  
OTHER INFORMATION: n equals a, t, g or c  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (145171)..(145171)  
OTHER INFORMATION: n equals a, t, g or c  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (145942)..(145942)  
OTHER INFORMATION: n equals a, t, g or c  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (147197)..(147197)  
OTHER INFORMATION: n equals a, t, g or c  
FEATURE:

US-09-701-229-2 (1-448) x US-09-790-988-1 (1-640681)



Qy	11	IleValValGlyLeuGlyLysSerGlyMetSerLeuValArgTyrLeuAlaArgGly	30
Db	241263	TTAAATTTAGCAATGGGTTTAAACAGGATATCTGCATTAATTTTTTTTTTAAAAAAGGA	241204
Qy	31	LeuProPheAlaValAlaValAspThrArgGluAsnProProGluLeuAlaThrLeuArgAla	50
Db		::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::	
Db	241203	ATAAAACCTAAATATTATGATGAATCTAAACATCTCTCAAAATTTTATTTAAATTT	241150
Qy	51	GlnTyrProGln---ValGluValArgCysGlyGluLeuAspAlaGluPheLeuCysSer	69
Db	241149	-----CCTCAAAATATTGAATATAGTTAGGAAGTTTAGACCATCAGTGGATTTCTAGAA	241096
Qy	70	AlaArgGluLeuTyrValSerProGlyLeuSerLeuArgThrProAlaLeuValGlnAla	89
Db	241095	TCAGACTTAATTTGTTATTAAGTCGGGTATTTCTTCATTTAAACCTATTTTAAATTAAGCA	241036
Qy	90	AlaAlaGlyGlyValArgIleSerGlyAspIleAspLeuPheAlaArgGluAlaLysAla	109
Db	241035	CGTTTGCTAGTATTGAGATTATTAGTGTATATTGAACATATTCTTAGAGAAGTACCTGT	240976
Qy	110	ProIleValAlaIleThrGlySerAsnAlaLysSerThrValThrThrLeuValGlyGlu	129
Db	240975	CCTATTATTCAATAACAGGACACTTGAAGAAAAGTACTGTGCTACCATGATTCAGAAAA	240916
Qy	130	MetaValAlaAlaAspLysArgValAlaValGlyGlyAsnLeuGlyThrProAlaLeu	149
Db	240915	ATTGCCAAAAAATCAGGATATATAAGCTTTTCTAGTGCTAATATAGGTGTTCCCGTACTA	240856
Qy	150	AspLeuLeuAlaAspAspIleGluLeuTyrValLeuGluLeuSerSerPheGlnLeuGlu	169
Db	240855	GAATACTCTGATAAAGAACGACAGATTATACATAATAGAACCTGTCTAGTTTCAACTAGAA	240796
Qy	170	ThrCysAspArgLeuAsnAlaGluValAlaThrValLeuAsnValSerGluAspHisMet	189
Db	240795	AATACATTTAATTTAAATCTACAGTAGCAGTTATCTTAATATTAGCGACAGATCATATC	240736
Qy	190	AspArgTyr---AspGlyMetaAspTyrHisLeuAlaLysHisArgIlePheArgGly	208
Db	240735	AATCGATACCCAAATGGATTCCACAATATATAAACACTAAATTTGCTGTTTATAATCAA	240676
Qy	209	AlaArgGlnValValValAsnArgAlaAspAlaLeuThrArgProLeuIleAlaAspThr	228
Db	240675	GCAGAAATTTGTATAAATTAATCTAATGATAGATAGAAAAGACTTTATTTCATAGTAAG	240616
Qy	229	ValProCysTrp---SerPheGlyLeuAsnLysProAspPheLysAlaPheGlyLeuIle	247
Db	240615	AATAAAAAATGGATTAGTTTGGACAATAAGAGTGAATTCGT-----ATTGTT	240565
Qy	248	GluGluAspGlyGlnLysTrpLeuAlaPheGlnPheAspLysLeuLeuProValGlyGlu	267
Db	240564	TCATAAGCAATGACCCCTATTATTTTAAAAATAAAAAATAATAATAACTAGTGAA	240505
Qy	268	LeuLysIleArgGlyAlaHisAsnTyrSerAsnAlaLeuAlaAlaLeuAlaLeuGlyHis	287
Db	240504	ATATTATTATGATATCAATAATACATAATATATATAGTCTCTTTAGCTATTTTCAGAT	240445
Qy	288	AlaValGlyLeuProPheAspAlaMetaMetLeuGlyAlaLeuLysAlaPheSerGlyLeuAla	307
Db	240444	GCAATGCAATTCCTAGNAATGACGCAATAATGTACTTAAAGTTTCTCGAATTTACCA	240385
Qy	308	HisArgCysGlnTrpValArgGluArgGlnGlyValSerTyrTyrAspAspSerLysAla	327
Db	240384	CATCGATTTCAATAATAAAAAATCAAAAAGGAGTACGTTGGATAAAGCATTTCTAAATCC	240325
Qy	328	ThrAsnValGlyAlaAlaLeuAlaAlaIleGluGlyLeuGlyAlaAspIleAspGlyLys	347
Db	240324	ACAAATGTAAATAGTACTCAAGTTGCTTTAAATTTCTATA-----AAAACAACTGGAACA	240271
Qy	348	LeuValLeuLeuAlaGlyGlyAspGlyLysGlyAlaAspPheHisAspLeuArgGluPro	367
Db	240270	ATAAGATTTATTATAGCGCGGATAGTAATCTGCAAATTTTAAATATATATATATATAT	240211
Qy	368	ValAlaArgPheCysArgAlaValValLeuLeuGlyArgAspAlaGlyLeuIleAlaGln	387

```

Db      240210 TTTAGAACTTTAAAAAATAAAATTTACTGTTTGAAGAGATGGCAATTAATTAATGTCAGAAA 240151
Qy      388 AlaLeuGlyAsnAlaValProLeuValArgValAlaThrLeuAspGluAlaValArgGln 407
Db      240150 ATATGTGAAAAA-----AATCTATTATGTTGAATAATTTAAAAAACCACTAATTTTA 240097
Qy      408 AlaAlaGluLeuAlaAraGluGlyAspAlaValLeuLeuSerProAlaCysAlaSerLeu 427
Db      240096 ATTTCAAAACAAGTCAAAATCAGGTGATACCGTTCTTTCTCTCTGGGTGCAGCAGCTTA 240037
Qy      428 AspMetPheLysAsnPheGluGluArgGlyArgLeuPheAlaLysAlaValGluGluLeu 447
Db      240036 GGTCAATTTTCTAACTTTGAAGAGAGAGTAATCTTTTATATAAAATTAATAAGGAAATA 239977

RESULT 8
US-10-156-761-6097
; Sequence 6097, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 6097
; LENGTH: 1401
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1401)
US-10-156-761-6097

Alignment Scores:
Pred. No.: 2,14e-63 Length: 1401
Score: 618.50 Matches: 180
Percent Similarity: 50.21% Conservative: 55
Best Local Similarity: 38.46% Mismatches: 188
Query Match: 27.37% Indels: 45
DB: 14 Gaps: 14

US-09-701-229-2 (1-448) x US-10-156-761-6097 (1-1401)
Qy      12 ValValGlyLeuGlyLysSerGlyMetSerLeuValArgTyrLeuAlaArgArgGlyLeu 31
Db      31 GTCGCGGCGTCTCGGCGTCTCCGGTATCCGGCGCGCGCGTCTGCACGCGCTCGCGCGG 90
Qy      32 PropheAlaValVal-----AspThrArgGluAsnProProGluLeuAlaThrLeu 48
Db      91 GTCGTCACGGTCGTCAAGCAGCGCGCACGACGAGCGTCCCGGCGCCCGACGCCGCGGATCTG 150
Qy      49 ArgAlaGlnTyrProGlnValGluValArgCysGlyGluLeuAspAlaGluPheLeuCys 68
Db      151 GAGGCGCTC-----GGCATTCACCGTCCGCTCGGT-----GACGGCGCGACCTCGCC 198
Qy      69 SerAlaArgGluLeuTyrVal---SerProGlyLeuSerLeuArgThrProAlaLeuVal 87
Db      199 GAGGCGACGAGCTCATGCTACACACCCGCTGGCAGCGGACGAGCGCGTGTTCGCC 258
Qy      88 GlnAlaAlaLysGlyValArgTlleSerGlyAspIleAspLeuPheAlaArg----- 105

```

Db 259 GCGGCGCGGAGCGCGCTCCCGCTGCTGGGCGGACGTGCACTCGCCCTGGCGGCTGGCC 318  
QY 106 ---GluAlaLysAlaProIleValAlaIleThrGlySerAsnAlaLysSerThrValThr 124  
Db 319 GCGCGCGGATCCGCGCCCTGGCTGCGCTCACGGGCAACCAAGGCAACACCGACCGCTG 378  
QY 125 ThrLeuValGlyGluMetAlaValAlaAlaAspLysArgValAlaValGlyGlyAsnLeu 144  
Db 379 CAGATGCTCGCTCGATCGATGCGCGCGCGGCGCTCGCACGCGCGCTGGCGCAACATC 438  
QY 145 GlyThrProAlaLeuAsp-----LeuLeuAlaAspAspIleGluLeuThrValLeuGlu 162  
Db 439 GCGGTCTCGCTCTCGACCGGCTCTCGCGGAGGACGTACGACGCTCTCGCGTGCAGAA 498  
QY 163 LeuSerSerPheGlnLeuGluThrCysAspArgLeuAsnAlaGluValAlaThrValLeu 182  
Db 499 CTCCTCAGCTACGAGCTGCTGCTGCGGCGCTCCCTGCGCGCCACATCCGCCACCGCTCG 558  
QY 183 AsnValSerGluAspHisMetAspArgTyrAspGlyMetAlaAspTyrHisLeuAlaLys 202  
Db 559 ACATCGCGCGGACGACCTCGACTGGCAGCGCTCCATGGAGGCGGTACACCGCGGACAG 618  
QY 203 HisArgIlePheArgGlyAlaArg---GlnValValAlaAspArgAlaAspAlaLeuThr 221  
Db 619 GCGGTATCTACGAGGCAATCGGTGCGCTCGCTACAGTGGCGGCGCTCGCTTCACGCC 678  
QY 222 ArgProLeuIle-----AlaAspThrValPro---CysTrpSerPheGly-----Leu 236  
Db 679 GAGGACCTGTGTCGCGCGGACGTCGAGAGGCGTGGCGGCGCTCGCTTCACGCC 738  
QY 237 AsnLysProAspPheLysAlaPheGlyLeuIleGlu-----248  
Db 739 GGCACGCGCGGACGTCCTCACTCGCGCTGCTGGAGGCGATCTCTGTCGACCGCGCTTC 798  
QY 249 ---GluAspGlyGlnLysTrpLeuAlaPheGlnPheAspLysLeuLeuProValGlyGlu 267  
Db 799 GTCGAGGACCGGCAAG-----ACGCGCGGAGCTGCGCGAGTGCGCGAC 846  
QY 268 LeuLysIleArgGlyAlaHisAsnTyrSerAsnAlaLeuAlaLeuAlaLeuGlyHis 287  
Db 847 GTGATCCCGCGCGCGCGCACACATCGCCACGCGCTTGGCGGCGCGCGCTGGCGCGC 906  
QY 288 AlaValGlyLeuProPheAspAlaMetLeuGlyAlaLeuLysAlaPheSerGlyLeuAla 307  
Db 907 GCTTCGCGGCTGCGCGCTGCGGCTGCGGAGGAGTGGCGGCTTGGCGCGCGCGCGC 966  
QY 308 HisArgCysGlnTrpValArgGluArgGlnGlyValSerTyrTyrAspSerLysAla 327  
Db 967 CACCGCATCGCGCGCTGCGCGCGCGCGCGCGCGCTGCGCGCGCGCGCGCGCGCGC 1026  
QY 328 ThrAsnValGlyAlaAlaLeuAlaIleGluGlyLeuGlyAlaAspIleAspGlyLys 347  
Db 1027 ACCAACACGCGCGCGGAGCTCTTTGGCGGCTGCGGCTGCGCGCGCGCGCGCGC 1071  
QY 348 LeuValLeuLeuAlaGlyAspGlyLysGlyAlaAspPheHisAspLeuArgGluPro 367  
Db 1072 ATCGCTGATGCGCGCGGCGCTGCGCAAGGCGCTCTCTGCGAGCTGCTGCGCGCAAG 1131  
QY 368 ValAlaIleArgPheCysArgAlaValValLeuLeuGlyArgAspAlaGlyLeuAlaGln 387  
Db 1132 TCGGCACAGCGCTCGCGGCTGTGTGTCTATCGCTGCGGATCGTGGCTGATCGTGA 1191  
QY 388 AlaLeuGlyAsnAlaValProLeuValArgValAlaThrLeuAspGlu-----403  
Db 1192 GCGCTTGGCGGACACGCGCGGAGGTACCGCTGCTGCGCTGCGCGCGGACGACACTGG 1251  
QY 404 -----AlaValArgGlnAlaGluLeuAlaArgGluGlyAspAlaValLeu 419  
Db 1252 GCGATCGCGCGCTGCTCAGAGGCGCGGCGCTGCGCGCTGCGCGCGGAGCAGCTACTG 1311  
QY 420 LeuSerProAlaCysAlaSerLeuAspMetPheLysAsnPheGluGluGlyArgLeu 439  
Db 1312 CTGGCGCGCGCTGTGCTCGATGACATGTTCGCCACTACACACGCGCGGCTGACCGC 1371

QY 440 PheAlaLysAlaValGluLeu 447  
Db 1372 TTCGCGGAGCGGTTTCGCGGATC 1395

## RESULT 9

US-10-156-761-1  
; Sequence 1, Application US/10156761  
; Publication No. US20030119018A1  
; GENERAL INFORMATION:  
; APPLICANT: OMURA, SATOSHI  
; APPLICANT: IKEDA, HARUO  
; APPLICANT: ISHIKAWA, JUN  
; APPLICANT: HORIKAWA, HIROSHI  
; APPLICANT: SHIBA, TADAYOSHI  
; APPLICANT: SAKAKI, YOSHIYUKI  
; APPLICANT: HATTORI, MASAHIRA  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-262  
; CURRENT APPLICATION NUMBER: US/10/156,761  
; CURRENT FILING DATE: 2002-05-29  
; PRIOR APPLICATION NUMBER: JP 2001-204089  
; PRIOR FILING DATE: 2001-05-30  
; PRIOR APPLICATION NUMBER: JP 2001-272697  
; PRIOR FILING DATE: 2001-08-02  
; NUMBER OF SEQ ID NOS: 15109  
; SEQ ID NO 1  
; LENGTH: 9025608  
; TYPE: DNA  
; ORGANISM: Streptomyces avermitilis  
; FEATURE:  
; NAME/KEY: misc.feature  
; LOCATION: (4187715)  
; OTHER INFORMATION: a, t, c, g, other or unknown  
US-10-156-761-1

Alignment Scores:  
Pred. No.: 7,39e-58 Length: 9025608  
Score: 618.50 Matches: 180  
Percent Similarity: 50.21% Conservative: 55  
Best Local Similarity: 38.46% Mismatches: 188  
Query Match: 27.37% Indels: 45  
DB: 14 Gaps: 14

US-09-701-229-2 (1-448) x US-10-156-761-1 (1-9025608)

QY 12 ValValGlyLeuGlyLysSerGlyMetSerLeuValArgTyrLeuAlaArgArgGlyLeu 31  
Db 7360919 GTCGCGGCGCTCGCGCTCTCCGCTATCCGCGCGCGCGCGCTGCTGACGCGCTCGCGCGG 7360978  
QY 32 ProPheAlaValVal-----AspThrArgGluAsnProProGluLeuAlaThrLeu 48  
Db 7360979 GTCGTCACGGTCTCAACGACGCGGCGGAGGCTCCCGGCGCGCGCGCGCGCGCTGCTG 7361038  
QY 49 ArgAlaGlnTyrProGlnValGluValArgCysGlyGluLeuAspAlaGluPheLeuCys 68  
Db 7361039 GAGGCGCTC-----GGCATCACCGCTGCGCTCGGT-----GACGCGCGCGCGCGCTG 7361086  
QY 69 SerAlaArgGluLeuTyrVal---SerProGlyLeuSerLeuArgThrProAlaLeuVal 87  
Db 7361087 GAGGCGACCGAGCTCATCTGTCACACCGCGCTGCGGAGCGCGGAGCGCGCGCTGCTGTCGC 7361146  
QY 88 GlnAlaAlaLysGlyValArgIleSerGlyAspIleAspLeuPheAlaArg-----105  
Db 7361147 GCGGCGCGCGGCGCGCGCTCCCGCTGCTGGGCGAGCTCGAAGCTCGCTGCGCGCTGCGCG 7361206  
QY 106 ---GluAlaLysAlaProIleValAlaIleThrGlySerAsnAlaLysSerThrValThr 124  
Db 7361207 GCGCGCGGATCGCGCGCTGCTGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTG 7361266  
QY 125 ThrLeuValGlyGluMetAlaValAlaAlaAspLysArgValAlaValGlyGlyAsnLeu 144  
Db 7361267 CAGATGCTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 7361326

```
Oy 145 GlyThrProAlaLeuAsp-----LeuLeuAlaAspIleGluLeuValLeuGlu 162
||| ||||| ||| : : : : : : : : : : : : : : : : : : : :
Db 7361327 GGCCTCTCGCTCTCGACGGGTCTCTCGCGAGGAGCTACGACGCTCTCGCGGTGAA 7361386

Oy 163 LeuSerSerPheGlnLeuGluThrCysAspArgLeuAsnAlaGluValAlaThrValLeu 182
||||| : : : : : : : : : : : : : : : : : : : :
Db 7361387 CTCTCCAGCTACCACTGCTGCGGCGCTCTCGCGCGCCACCTCCGCGACGCTCTG 7361446

Oy 183 AsnValSerGluAspHisMetAspArgTyrAspGlyMetAlaAspTyrHisLeuAlaLys 202
||| : : : : : : : : : : : : : : : : : : : :
Db 7361447 AACATCGCGCGGACCTCGACTGCGCTGCGGCTTCCGAGGGGTACACCGCGGCAAG 7361506

Oy 203 HisArgIlePheArgGlyAlaArg---GlnValValAsnArgAlaAspAlaLeuThr 221
||||| : : : : : : : : : : : : : : : : : : : :
Db 7361507 GGCGGTATCTACGAGGCAATCGGGTCTGCTCTACAACTGGCGCGCAAGGCCACC 7361566

Oy 222 ArgProLeuIle-----AlaAspThrValPro---CysTrpSerPheGly-----Leu 236
||||| ||||| ||||| : : : : : : : : : : : : : :
Db 7361567 GAGGACCTGTGTGCGCGCGGAGCTGCGAGAGGGTCTCGCGCGCTTACGCTG 7361626

Oy 237 AsnLysProAspPheLysAlaPheGlyLeuIleGlu-----248
||| : : : : : : : : : : : : : : : : : : : :
Db 7361627 GGCAGCGCGGACCGTCCCACTCGGCTGCTGAGGGCATCTGTGTCGACCGCGCTTC 7361686

Oy 249 ---GluAspGlyGlnLysTrpLeuAlaPheGlnPheAspLysLeuLeuProValGlyGlu 267
||||| ||||| ||||| : : : : : : : : : : : : : :
Db 7361687 GTCGAGGACCGGCAAG-----AACGCCAGGAGCTGCGCGAGGTGGCGGAC 7361734

Oy 268 LeuLysIleArgGlyAlaHisAsnTyrSerAsnAlaLeuAlaLeuAlaLeuGlyHis 287
||| : : : : : : : : : : : : : : : : : : : :
Db 7361735 GTGCATCGCGCGCGCGGCAACATCCCAACGCCCTTGGCGAGCGCGCTGGCGCGC 7361794

Oy 288 AlaValGlyLeuProPheAspAlaMetLeuGlyAlaLeuLysAlaPheSerGlyLeuAla 307
||| ||||| : : : : : : : : : : : : : : : : : : : :
Db 7361795 GCCTTCGGGTGCGCGCTCGCGGTACGCGGAGCTGCGGCTTCGCGCGGAGGCC 7361854

Oy 308 HisArgCysGlnTrpValArgGluArgGlnGlyValSerTyrTyrAspSerLysAla 327
||||| ||||| : : : : : : : : : : : : : : : : : : : :
Db 7361855 CACGCGATCGCGCAGCTGCGCGAGCTGCGCGGCTACCTACATCGAGCTCCCAAGGCC 7361914

Oy 328 ThrAsnValGlyAlaAlaLeuAlaIleGluGlyLeuGlyAlaAspIleAspGlyLys 347
||||| ||||| ||||| : : : : : : : : : : : : : :
Db 7361915 ACCAAGCAGCGCGCGGAGCCTCTTGGCGCGCTACGCGGTCC-----7361959

Oy 348 LeuValLeuLeuAlaGlyAspGlyLysGlyAlaAspPheHisAspLeuArgGluPro 367
||| : : : : : : : : : : : : : : : : : : : :
Db 7361960 ATCGTCTGGATCGCGCGGCTCGCCAGGCGCTCTCTCGACGAGCTGGTGGCCAG 7362019

Oy 368 ValAlaArgPheCysArgAlaValValLeuLeuGlyArgAspAlaGlyLeuIleAlaGln 387
||||| : : : : : : : : : : : : : : : : : : : :
Db 7362020 TCGGCACAGCGGCTCCGGGTGTGTGCTCATCGTGGTGGCTGATCCGTGAA 7362079

Oy 388 AlaLeuGlyAsnAlaValProLeuValArgValAlaThrLeuAspGlu-----403
||||| ||||| ||||| : : : : : : : : : : : : : :
Db 7362080 GCCCTTGGCGACACGCGCGGAGGTACCCGCTGCGACCTCGACCGCGGACACACATGG 7362139

Oy 404 -----AlaValArgGlnAlaAlaGluLeuAlaArgGluGlyAspAlaValLeu 419
||||| : : : : : : : : : : : : : : : : : : : :
Db 7362140 CGGATGCGCGGCTGTCCAGGAGCGCGGCTCGCGCTCGCGCGGACACGCTACTG 7362199

Oy 420 LeuSerProLacCysAlaSerLeuAspMetPheLysAsnPheGluArgGlyArgLeu 439
||||| ||||| : : : : : : : : : : : : : : : : : : : :
Db 7362200 CTGCGCGCGGCTGTGCTCGATGCATGTTCGCCCAATACCAACAGCGCGGTGACGCG 7362259

Oy 440 PheAlaLysAlaValGluGluLeu 447
||||| : : : : : : : : : : : : : : : : : : : :
Db 7362260 TTCGCGAGCGGTTTCGCGGACTC 7362283

RESULT 10
US-09-815-242-9535
; Sequence 9535, Application US/09815242
; Patent No. US20020061569A1
```

```
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR FILING DATE: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9535
; LENGTH: 1353
; TYPE: DNA
; ORGANISM: Streptococcus pneumoniae
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1353)
; US-09-815-242-9535

Alignment Scores:
Pred. No.: 2,49e-60 Length: 1353
Score: 592.50 Matches: 163
Percent Similarity: 52.92% Conservative: 82
Best Local Similarity: 35.21% Mismatches: 177
Query Match: 26.22% Indels: 41
DB: 9 Gaps: 14

US-09-701-229-2 (1-448) x US-09-815-242-9535 (1-1353)

Oy 7 AsphisPheArg-----IleValValGlyLeuGlyLysSerGlyMetSerLeu 22
||| ||||| : : : : : : : : : : : : : :
Db 13 GATCAATTTAAATAAGAAAGTTCTTGTGTTAGTTTGGCAAGTCTGGTGAATCTGCG 72

Oy 23 ValArgTyrLeuAlaArgArgGlyLeuProPheAlaValValAspThrArg-----39
||| ||||| : : : : : : : : : : : : : :
Db 73 GCTCGTTGTGGACAAGTTAGGTGCCATTGTGCACAGTAAATGATGGAAACCATTCGAG 132

Oy 40 GluAsnProProGluLeuAlaThrLeuArgAlaGlnTyrProGlnValGluValArgCys 59
||| ||||| : : : : : : : : : : : : : :
Db 133 GACAAATCCAGTCCCAATGTTTCTGCTGAA-----GAAGGATCAAGGTCAATTACA 183

Oy 60 GlyGlu-----LeuAspAlaGluPheLeuCysSerAlaArgGluLeuTyr 74
||| ||||| : : : : : : : : : : : : : :
Db 184 GGTGCCATCCTTTTGAACATCTTGTGATGAAGAGTTGCCCCATCGTTAATAA-----234

Oy 75 ValSerProGlyLeuSerLeuArgThrProAlaLeuValGlnAlaAlaLysGlyVal 94
||| ||||| : : : : : : : : : : : : : :
Db 235 ---AATCAGGTATCCCTACAGCAATCCCATGATTGAAAGGCTTTGGCCCAAGGGAAT 291

Oy 95 ArgIleSerGlyAspIleAspLeuPheAlaArgGluAlaLysAlaProIleValAlaIle 114
||| : : : : : : : : : : : : : : : : : : : :
Db 292 CCAGTCTTGACTGAGTGAATGCTTATTTGATTTCAGAAGCACCGATTTATTGGTATC 351
```

```
QY 115 ThrGlySerAsnAlaValGlyGlySerThrValThrLeuValGlyGluMetAlaValAla 134
Db 352 ACAGGATCGAAGCGTGAAGCAACACACAGCATATGATGGGAAGCTTTTGACTGCTGCT 411
QY 135 AsPlysArgValAlaValGlyGlyAsnLeuGlyThrProAlaLeuAspLeu-----Leu 152
Db 412 GGGCAACATGGCTTTTATCAGGGAATATCGGCTATCCTGCCAGTCAGGTGCTCAAAAT 471
QY 153 AlaAspAspIleGluLeuTyrValLeuGluLeuSerSerPheGlnLeuGluThrCysAsp 172
Db 472 GCATCAGATAAGGACACGCTGTTGATGGAATTTCTTCTTCCAACTCATGGGTGTTCAA 531
QY 173 ArgLeuAsnAlaGluValAlaThrValLeuAsnValSerGluAspHisMetAspArgTyr 192
Db 532 GAATTCATCCAGAGATTCGGTTATTAACCACTCATGCCAACTCATATCGACTACCAT 591
QY 193 AspGlyMetAlaAspTyrHisLeuAlaLysHisArgIle-----PheArgGlyAla 209
Db 592 GGGTCATTCTTGAATATAGCAGCAAGTGGATATCCAGAACAGATGACAGCAGCT 651
QY 210 ArgGlnValValAsnArgAlaAspAlaLeuThrArgProIleAlaAspThr---- 228
Db 652 GATTCCTGCTCTTGAATTAATCAAGACTTGGCAAAAGACTTGACTTCCAAGACAGAA 711
QY 229 -----ValProCysTrpSerPheGlyLeuAsnLysProAspPheLysAlaPheGly 245
Db 712 GCACATGTTGTACCAATTTTCAACA-----CTGAAAGGTTGAT-----GGA 753
QY 246 LeuIleGluGluAspGlyGlnLysTrpLeuAlaPheGlnPheAspLysLeuLeuProVal 265
Db 754 GCTTATCTGGAGATGGTCAA-----CTCTACTCCGTTGGTGAAGTAGTCAATGCGAGCG 807
QY 266 GlyGluLeuLysIleArgGlyAlaHisAsnTyrSerAsnAlaLeuAlaAlaLeu 285
Db 808 AATGAATCGGTGTTCCAGGTAGCCACAATGTGAAATGCCCTTGGCGACTATTGCTGTA 867
QY 286 GlyHisAlaValGlyLeuProPheAspAlaMetLeuGlyAlaLeuLysAlaPheSerGly 305
Db 868 GCCAGCTCGTGATGTGGACAAATCAACCACTCAAGAAACCTTTTCAGCCCTTCGGTGT 927
QY 306 LeuAlaHisArgCysGlnTrpValArgGluArgGlnGlyValSerTyrTyrAspAspSer 325
Db 928 GTCAAAACACCGTCTCCAGTTTGTGGATGACATCAAGGGTGTAAATCTTATACACAGCT 987
QY 326 LysAlaThrAsnValGlyAlaAlaLeuAlaIleGluGlyLeuGlyAlaAspIleAsp 345
Db 988 AAATCAACTAATATCTTGGCTACTCAAAAGCCCTTATCAGGATTTAC-----AAC 1038
QY 346 GlyLysLeuValLeuAlaGlyGlyAspGlyLysGlyAlaAspPheHisAspLeuArg 365
Db 1039 ACCAGGTGCTGTTGATGTCAGGTGTTTGGACCGTGGCAATGAGTTGACGAATG--- 1095
QY 366 GluProValAlaArgPheCysArgAlaValValLeuLeuGlyArgAspAlaGlyLeuIle 385
Db 1096 GTGCCAGACATTAATCTGAGTCAAGAGATGTTGCTATCCTGGGTCAATCTGCAGACGTGTC 1155
QY 386 AlaGlnAlaLeuGlyLysAsnAla---ValProLeuValArgValAlaThrLeuAspGluAla 404
Db 1156 AAACGGGCGACAGACAGCGCTGGTGTGCTTATGTGGAGGCGACAGATATTGCAGATGCG 1215
QY 405 ValArgGlnAlaAlaArgGluGlyAspAlaValValLeuLeuSerProAlaCys 424
Db 1216 ACCCGCAGGCTATGACGTGCGACTCAAGAGATGTTGTTCTTCTTAGTCCCTGCCAAT 1275
QY 425 AlaSerLeuAspMetPheLysAsnPheGluGluArgGlyArgLeuPheAlaLysAlaVal 444
Db 1276 GCCAGCTGGATATGATGCTAACTTTGAAGTACGTGGCGACCTCTTCTATCGACACAGTA 1335
QY 445 GluGluLeu 447
Db 1336 GCGGAGTTA 1344
RESULT 11
```

```
US-09-815-242-9300
; Sequence 9300, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9300
; LENGTH: 1353
; TYPE: DNA
; ORGANISM: Streptococcus pneumoniae
; FEATURE:
; LOCATION: (1)...(1353)
; US-09-815-242-9300

Alignment Scores:
Pred. No.: 1,298-59 Length: 1353
Score: 586.50 Matches: 162
Percent Similarity: 52.92% Conservative: 83
Best Local Similarity: 34.95% Mismatches: 177
Query Match: 25.95% Indels: 41
DB: 9 Gaps: 14

US-09-701-229-2 (1-448) x US-09-815-242-9300 (1-1353)

QY 7 AspHisPheArg-----IleValValGlyLeuGlyLysSerGlyMetSerLeu 22
Db 13 GATCAATTAAAAATAAGAAAGTCTTGTTTTAGTTGGCCCAAGTCTGGTGAATCTGCA 72
QY 23 ValArgTyrLeuAlaArgArgGlyLeuProPheAlaValValAspThrArg----- 39
Db 73 GCTCGTTTGTGGCAAGCTAGTGCCATTGTGACAGTAAATGATGGGAAACCTTTTCGAG 132
QY 40 GluAsnProGluLeuAlaThrLeuArgAlaGlnTyrProGlnValGluValArgCys 59
Db 133 GACATCCAGCTGCCAAAGTTTGTGGAA-----GAAGGATCAAGGTCTATACA 183
QY 60 GlyGlu-----LeuAspAlaGluPheLeuCysSerAlaArgGluLeuTyr 74
Db 184 GGTGCCATCTCTTGGAACTCTTGGATGAAGAGTTTGCCTTATGGTGAAG----- 234
QY 75 ValSerProGlyLeuSerLeuArgThrProAlaLeuValGlnAlaAlaLysGlyVal 94
Db 235 ---AATCCAGGTATCCCTTACAAACAATCCCATGATTGAAAGGCTTTGCCCAAGAGAAT 291
QY 95 ArgIleSerGlyAspIleAspLeuPheAlaArgGluAlaLysAlaProIleValAlaIle 114
```

```
Db 292 CCAGTCTTGAGTGGAGTGGCTATTGTTGATTCAGAAGCACCGGATTATTGGTATC 351
Qy 115 ThrGlySerAsnAlaLysSerThrValThrThrLeuValGlyGluMetAlaValAla 134
Db 352 ACAGATCGAAGCGTAGAACACACACAGCACTATGATGGGGAAGTTTGTAGCTGCT 411
Qy 135 AspLysArgValAlaValGlyGlyAsnLeuGlyThrProAlaLeuAspLeu-----Leu 152
Db 412 GGGCAACATGCTTTTATCATCGGAATATCGCTATCTCCGACCTCAGGTGCTCAATA 471
Qy 153 AlaAspIleGluLeuThrValLeuGluLeuSerPheGlnLeuGluThrCysAsp 172
Db 472 GCATCAGATAGGACAGCGCTTGTATGAACTTCTTCTTCCAACTCATCGGTGTCAA 531
Qy 173 ArgLeuAsnAlaGluValAlaThrValLeuAsnValSerGluAspHisMetAspArgTyr 192
Db 532 GAATTCATCAGAGATGGCGTTATACCAACTCATGCCAATCATATCGATCATCAT 591
Qy 193 AspGlyMetAlaAspTyrHisLeuAlaLysHisArgIle-----PheArgGlyAla 209
Db 592 GGGTCATTTCTGAATATGTAGCAGCAAGTGAATATCCAGAACAGATGACAGCAGCT 651
Qy 210 ArgGlnValValAlaAsnArgAlaAspAlaLeuThrArgProLeuIleAlaAspThr--- 228
Db 652 GATTCTCTGCTTGAACCTTAAATCAAGACTTGCACAAAGACTTGACTTCCAAAGACAA 711
Qy 229 -----ValProCysTrpSerPheGlyLeuAsnLysProAspPheLysAlaPheGly 245
Db 712 GCCACTGTGTACCATTTTCAACA-----CTTGAAGAAGTTGAT-----GGA 753
Qy 246 LeuIleGluAspGlyGlnLysTrpLeuAlaPheGlnPheAspLysLeuLeuProVal 265
Db 754 GCTTATCTGGAAGATGGTCAA-----CTCTACTTCGCTGGTGAAGTAGTATGCGCAGCG 807
Qy 266 GlyGluLeuLysIleArgGlyAlaHisAsnTyrSerAsnAlaLeuAlaAlaLeuAlaLeu 285
Db 808 AATGAATCGGTGTCCAGGTAGGCACCAATGTGAAATGCCCTGGGACATTTGCTGTA 867
Qy 286 GlyHisAlaValGlyLeuProPheAspAlaMetLeuGlyAlaLeuLysAlaPheSerGly 305
Db 868 GCCAAGCTCTCGTGTGGCAATCAACCATCAAGGAACCTCTTCAGCTTCGTTGGT 927
Qy 306 LeuAlaHisArgCysGlnTrpValArgGluArgGlnGlyValSerTyrTyrAspSer 325
Db 928 GTCAAAACACCGCTCCAGCTTGTGATGCATCAAGGGTGTAAATTTCTATAACGACAGT 987
Qy 326 LysAlaThrAsnValGlyAlaAlaLeuAlaIleGluGlyLeuGlyAlaAspIleAsp 345
Db 988 AAATCAACTAATATCTTGGCTACTCAAAAGCCCTATCAGGATTTGAC-----AAC 1038
Qy 346 GlyLeuValLeuLeuAlaGlyGlyAspGlyLysGlyAlaAspPheHisAspLeuArg 365
Db 1039 ACRAAGTCTGCTGATTCAGGTGGTGGTGGACCGTGGCAATGATTTGACGAATG--- 1095
Qy 366 GluProValAlaAlaArgPheCysArgAlaValValLeuLeuGlyArgAspAlaGlyLeuIle 385
Db 1096 GTGCCAGACATTTACTGGACTCAAGAAGATGCTCATCTCGGTCTCAATCTGCAGACGCTC 1155
Qy 386 AlaGlnAlaLeuGlyAsnAla---ValProLeuValArgValAlaThrLeuAspGluAla 404
Db 1156 AAAGGGGCGAGCAAGCGCTGTGCTTATGAGGGGCGACAGATATTCAGATGCG 1215
Qy 405 ValArgGlnAlaAlaGluLeuAlaArgGluGlyAspAlaValLeuLeuSerProAlaCys 424
Db 1216 ACCGCAAGGCTATGACTTGCAGCTCAAGAGATGTGGTCTTCTTAGTCTCCCAAT 1275
Qy 425 AlaSerLeuAspMetPheLysAsnPheGluGluArgGlyArgLeuPheAlaLysAlaVal 444
Db 1276 GCCAGCTGGGATATGATGCTAACTTTGAAGTACTGGGACCTCTTTATTCGACACAGTA 1335
Qy 445 GluLeuLeu 447
|||||
```

```
Db 1336 CGCGAGTTA 1344
RESULT 12
US-09-701-229-2 (1-1352)
; Sequence 1685, Application US/09974300
; Patent No. US20020146721A1
; GENERAL INFORMATION:
; APPLICANT: Berka, Randy M.
; TITLE OF INVENTION: Methods For Monitoring Multiple Gene
; TITLE OF INVENTION: Expression
; FILE REFERENCE: 10085-500-US
; CURRENT APPLICATION NUMBER: US/09/974,300
; CURRENT FILING DATE: 2001-10-05
; PRIOR APPLICATION NUMBER: 09/680,598
; PRIOR FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: 60/279,526
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 8481
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1685
; LENGTH: 1352
; TYPE: DNA
; ORGANISM: Bacillus licheniformis
US-09-701-229-2 (1-448) x US-09-701-229-2 (1-1352)
Alignment Scores:
Pred. No.: 1,92e-53 Length: 1352
Score: 534.50 Matches: 155
Percent Similarity: 50.32% Conservative: 78
Best Local Similarity: 33.48% Mismatches: 179
Query Match: 23.65% Indels: 51
DB: 10 Gaps: 16
US-09-701-229-2 (1-448) x US-09-701-229-2 (1-1352)
Qy 11 IleValValGlyLeuGlyLysSerGlyMetSer-----Leu 22
Db 37 CTCCTACTGGGGCTGGCGAAAGGGATATGCTCGCTGCTGATCTTCATGAGAAATGC 96
Qy 23 ValArgTyrLeuAlaArgArgGlyLeuProPheAlaValAlaValAspThrArgGluAsnPro 42
Db 97 GTAAAGGTGTCGTGAATGACGACGCTGCTTTT-----GAGGAAATGAA 141
Qy 43 ProGluLeuAlaThrLeuArgAlaGlnTyrProGlnValGluValArgCysGlyGluLeu 62
Db 142 CCT-----GCGCGATTCCTCGCGAA---AAAGGGTGGAGTCTGCGGCTCTCAT 192
Qy 63 AspAlaGluPhe-----LeuCysSerAlaArgGluLeuTyrValSerProGlyLeuSer 80
Db 193 CCGACGGAACTGTTTGTATCTTTCATTCGATCGACATTCGATTAATAATCCGGCATCCGC 252
Qy 81 LeuArgThrProAlaLeuValGlnAlaAlaLysGlyValArgIleSerGlyAspIle 100
Db 253 TATGAAACACGTGATGGTGCAGAGACGCTGAAGCGGGTATCCCGGCTCGACGAGGTT 312
Qy 101 AspLeuPheAlaArgGluAlaLysAlaProIleValAlaIleThrGlySerAsnAlaLys 120
Db 313 GAATCGGCTATCATCTGACAGATGCTCCGTTTCATCGGCATTCACGGTTCAACAGCAAG 372
Qy 121 SerThrValThrLeuValGlyGluMetAlaValAlaAlaAspLysArgValAlaVal 140
Db 373 ACACACGACACACGCTTGTATGAATGCTGAAGCGCGATTCGAAAGGCTCGCT 432
Qy 141 GlyGlyAsnLeuGlyThrProAlaLeuAspLeu-----AlaAspAspIleGluLeu 158
Db 433 GCTGGGAATATCGGTACGCGGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 492
Qy 159 TyrValLeuGluLeuSerSerPheGlnLeuGluThrCysAspArgLeuAsnAlaGluVal 178
Db 493 ATCTGTACAGAGCTGCTGAGTTTTCAGCTATGGGGAGCTTATCAATTCAGACCAAGAT 552
Qy 179 AlaThrValLeuAsnValSerGluAspHisMetAspArgTyrAspGlyMetAlaAspTyr 198
```

```
Db 553 GGATTATTAAACGATTGATGCCATCTCGATTATCACCATTCGCGCAAAATAT 612
QY 199 HisLeuAlaLysHisArgIlePheArgGlyAlaArgGln-----ValValValAsn 215
Db 613 GAGCTTGCANACAGCAGGTCTATCAATCAATCAATCAATCAATCAATCAATCAAT 672
QY 216 ArgAlaAspAlaLeuThrArgProLeuIleAlaAspThrValProCysTrpSerPheGly 235
Db 673 TTGGACGAT-----GAGACAGTCGTCGCGCTCGCGAGTGT-----708
QY 236 LeuAsnLysProAspPheLysAlaPhe-----GlyLeuIle 247
Db 709 ---TCAAAAGCGGAGAGGTGATTTTCGTCGCGCGGACACTTGAGCGCGGAGGTGC 765
QY 248 GluGluAspGlyGlnLysTrpLeuAlaPheGlnPheAspLysLeuLeuProValGlyGlu 267
Db 766 GTCAAGAGCGGCCA-----ATCATGTTTAAAGCGGACCGGTCAATCGGTGGAAGAC 819
QY 268 LeuLysIleArgGlyAlaHisAsnTySerAsnAlaLeuAlaAlaLeuAlaLeuGlyHis 287
Db 820 GTCGTTTTCGCGGAGAGCACAAATTTGGAGAACATTTTGGCGCCCTTTGCATCGTTAA 879
QY 288 AlaValGlyLeuProPheAspAlaMetLeuGlyAlaLeuLysAlaPheSerGlyLeuAla 307
Db 880 ACAGCAGCGGTTCGACGCGCGCTCGTAAAGTCTGACGAGCTTTACAGGTGTTAAG 939
QY 308 HisArgCysGlnTrpValArgGluArgGlnGlyValSerTyTrpAspSerLysAla 327
Db 940 CACAGGATCATACGTCCGCGAGTCAAAACAGACTGTTTACAAATCAGACGAAAGCG 999
QY 328 ThrAsnValGlyAlaAlaLeuAlaIleGluGlyLeuGlyAlaAspIleAspGlyLys 347
Db 1000 ACAACATT-----CTTCGCGAGAAAGCGCTGTCGCGC-----TTTCAAAAGCG 1047
QY 348 LeuValLeuLeuAlaGlyGlyAspGlyLysGlyAlaAspPheHisAspLeuArgLupro 367
Db 1048 GTCATTTTCTGCGAGGGGGCTGACCGCGGAAATGAATTTGATGAACATAAG---CG 1104
QY 368 ValAlaArgPheCysArgAlaValValLeuLeuGlyArgAspAla-----GlyLeu 384
Db 1105 CATATCTCTTTGTAAGCGGTGATCACTTCGCGGAGACCGCGCGGAGTTGAGAG 1164
QY 385 IleAlaGlnAlaLeuGlyAsnAlaValProLeuValArgValAlaThrLeuAspGluAla 404
Db 1165 CTGGCGGAGAAATGGGA---ATCAACAGGTAAACGTGTGATAATGTTGAACAAGCA 1221
QY 405 ValArgGlnAlaAlaGluLeuAlaArgGluGlyAspAlaValLeuLeuSerProAlaCys 424
Db 1222 GCAACTGCGCGCTTACGCTGTGACGACGAGGAGATGTCATCTCTGTCGCGCGCTGC 1281
QY 425 AlaSerLeuAspMetPheLysAsnPheGluGluArgGlyArgLeuPheAlaLysAlaVal 444
Db 1282 GCAAGCTGGGATCAGTACAAACATTTGAAGAACGTGGTGACATGTTTGTAAACCGCG 1341
QY 445 GluGluLeu 447
Db 1342 CATATGCTT 1350
```

## RESULT 13

```
US-09-712-363-79
; Sequence 79, Application US/09712363
; Patent No. US20020164588A1
; GENERAL INFORMATION:
; APPLICANT: Eisenberg, David
; APPLICANT: Rotstein, Sergio H.
; APPLICANT: Marcotte, Edward M.
; TITLE OF INVENTION: DETERMINING THE FUNCTIONS AND
; INTERACTIONS OF PROTEINS BY COMPARATIVE ANALYSIS
; FILE REFERENCE: 07419-032001
; CURRENT APPLICATION NUMBER: US/09/712,363
; CURRENT FILING DATE: 2000-11-13
; PRIOR APPLICATION NUMBER: PCT/US00/02246
```

```
; PRIOR FILING DATE: 2000-01-28
; PRIOR APPLICATION NUMBER: 60/179,531
; PRIOR FILING DATE: 2000-02-01
; PRIOR APPLICATION NUMBER: 60/117,844
; PRIOR FILING DATE: 1999-01-29
; PRIOR APPLICATION NUMBER: 60/118,206,
; PRIOR FILING DATE: 1999-02-01
; PRIOR APPLICATION NUMBER: 60/126,593
; PRIOR FILING DATE: 1999-03-26
; PRIOR APPLICATION NUMBER: 60/134,093
; PRIOR FILING DATE: 1999-05-14
; PRIOR APPLICATION NUMBER: 60/134,092
; PRIOR FILING DATE: 1999-05-14
; PRIOR APPLICATION NUMBER: 60/165,124
; PRIOR FILING DATE: 1999-11-12
; PRIOR APPLICATION NUMBER: 60/165,086
; PRIOR FILING DATE: 1999-11-12
; NUMBER OF SEQ ID NOS: 292
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 79
; LENGTH: 1461
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; US-09-712-363-79

Alignment Scores:
Pred. No.: 2,89e-52 Length: 1461
Score: 525.00 Matches: 167
Percent Similarity: 45.368 Conservative: 62
Best Local Similarity: 33.818 Mismatches: 182
Query Match: 23.23% Indels: 84
DB: 10 Gaps: 14
```

US-09-701-229-2 (1-448) x US-09-712-363-79 (1-1461)

```
QY 11 lIeValValGlyLeuGlyLysSerGlyMetSerLeuValArgTyTrpLeuAlaArgGly 30
Db 34 TTGGTAGCGGTTGGCGGCGGTGACCGGTCAAGCGGTGCGCGGTGCTGACTCGGTTTGT 93
QY 31 LeuProPheAlaValAlaAspThrArgGluAsnProGluLeuAlaThrLeuArgAla 50
Db 94 GCGACGCCACCGTGTGCGAC-----GACGATCCGGTCTGCTCGCA-----135
QY 51 GlnTyProGlnValGluValArgCysGlyGluLeuAspAlaGluPheLeuCysSer-- 69
Db 136 -----CCGACCGCGGACGCGGTGGGTGCTGCTCGGACCGGTGCGGCGGCGAG 189
QY 70 -----AlaArgGluLeuTyTrpValSerProGlyLeuSerLeuArgThrProAla 86
Db 190 ATAAACGGGTATGCGGTG--GTGGTCCGCTCCCGCTTCTCGCGGCAACCCGCTACT 248
QY 86 uValGlnAlaAlaLysGlyValArgIleSerGlyAspIleAspLeuPheAlaArgG 106
Db 249 GCGCGCGCGCGCGCGCGCGGTGCGGATCGCGGTGAGTGGAGTGTAGCTGCGGCGGT 308
QY 106 uAlaLysAlaProIle-----ValAlaIleThrGlySerAsnAl 119
Db 309 AGACGACGCGGCTGCTAGCGACCGCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTG 368
QY 119 aLysSerThrValThrLeuValGlyGluMetAlaValAlaAlaAspLysArgVal 139
Db 369 CAAGACCAACGACGATGATGTCACGCCATGTCGCGGTGCGGCGGTGCGGCGGCGGT 428
QY 139 aValGlyGlyAsnLeuGlyThrProAlaLeuAspLeuLeuAlaAspIleGluLeuTy 159
Db 429 GCTGTGCGGCATATCGCGAGTGGCGGTGCTGATGCTGGGACGAGCGCGCGGTGCT 488
QY 159 rValLeuGluLeuSerSerPheGlnLeuGluThrCysAspArgLeuAsnAlaGluVal 179
Db 489 GCGCGTGGAGTGTCCAGTTTCCAGTTCGACCTGCGGCGGCGGTGCGGCGGCGGCG 548
QY 179 aThrValLeuAsnValSerGluAspHisMetAspArgTyTrpAspGlyMetAlaAs 199
```

```
Db 549 CGCGGTGCTCAACATTCCGAGACACCTGGCTGATGCCACGATGCCGCAATACAC 608
Qy 199 sLeuAlaLysHisArgIlePheArgGlyAlaArgGlnValValAlaAspAl 219
Db 609 CGCGGCCAAGCCCGGTGTGACCGCGGG-----GTAGCGGTGGCCGGGTGATGA 662
Qy 219 aLeuThrArgProLeuIleAlaAspThrValProCysTrp-----SerPheGlyLe 236
Db 663 CAGCCGAGCGCGGCACTGTGACCGGCTCACCGCGCAGGTGGCGGTTCGGCT 722
Qy 236 uAsnLysProAspPheLysAlaPheGly-----LeuIleGluAspG1 251
Db 723 CGCGAGCCCGCGCGGGAATGGCGGTGGCGGACCCACCTGGTCGATCGC----- 777
Qy 251 yGlnLysTrpLeuAlaPheGlnPheAsp---LysLeuLeuProValGlyGluLeuLysI1 270
Db 778 -----GCCTTCTCCGACGACTTACCGCTGCTCCCGTCCGTCGATACCGGT 824
Qy 270 eArgGlyAlaHisAsnTyrSerAsnAlaLeuAlaLeuAlaLeuGlyHisAlaValG1 290
Db 825 GCCAGGTCCGCTGGCTGCTGACGCCCTGGCGCGCGCTGGCGCGCTCGGTGCG 884
Qy 290 yLeuProPheAspAlaMetLeuGlyAlaLeuLysAlaPheSerGlyLeuAlaHisArgCy 310
Db 885 GGTGCCCGCGGTGCGATGCCGCGGCTACCGCTGCTTCGAGTGGCGCGCACCGCGC 944
Qy 310 sGlnTrpValArgGluArgGlnGlyValSerTyrTyrAspSerLysAlaThrAsnVa 330
Db 945 CGAGGTGGTGGCGCTTCCCGACGCATCACCTACGTGGACGACTCCAGGCCACCAACC 1004
Qy 330 lGlyAlaAlaLeuAlaAlaIleGluGlyLeuGlyAlaAspIleAspGlyLysLeuValle 350
Db 1005 GCAGCGCGCGGTGCTGGCTGCTGCTGCTATAC-----CCGAGGTGGGTATG 1049
Qy 350 uLeuAlaGlyGlyAspGlyLysGlyAlaAspPheHisAspLeuArgGluProValAlaAr 370
Db 1050 GATCGCGCGGTGGCTGCTCAAGCGCGCTCGCTTCAC-----GCCAGGTGGCGCGAT 1103
Qy 370 gPheCysArgAlaVal-----ValLeuLeuGlyArgAspAlaGlyLeuIleAlaGlnAl 388
Db 1104 GCGTCCGCGCTGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1163
Qy 388 aLeuGlyAsn-----AlaValProLeuValArgVal----- 398
Db 1164 GTTATCACGACACGCGCGCGATGTCCAGTCTGCTCAGGTGTGTCGAGCGGAGGATAC 1223
Qy 398 ----- 398
Db 1224 TATGCTCGGACTGTGAGTTCCTGTTGCTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 1283
Qy 399 -----AlaThrLeuAspGluAlaValArgGlnAlaAlaGluLe 411
Db 1284 AGCGGTGAGACCGTGGCGTGGCGGTGATGACCGGTGGCGGTGGCGGTGGCGGTGG 1343
Qy 411 uAlaArgGluGlyAspAlaValLeuLeuSerProAlaCysAlaSerLeuAspMetPheLy 431
Db 1344 GGCCCAACCGGTGACACCGTGTGTCGGACCGCGCGCTCATTCGACCATTCAC 1403
Qy 431 sAsnPheGluGluArgGlyArgLeuPheAlaLysAlaVal 444
Db 1404 CGGTTATGCCGACCGCGCGGAGGCAATTCGCGACCGCGGT 1443
```

## RESULT 14

```
US-09-815-242-3936
; Sequence 3936, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlson, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
```

```
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: XU, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011a
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3936
; LENGTH: 1368
; TYPE: DNA
; ORGANISM: Enterococcus faecalis
US-09-815-242-3936
```

## Alignment Scores:

Pred. No.:	1.36e-51	Length:	1368
Score:	519.00	Matches:	152
Percent Similarity:	51.09%	Conservative:	83
Best Local Similarity:	33.04%	Mismatches:	181
Query Match:	22.96%	Indels:	44
DB:	9	Gaps:	16

US-09-701-229-2 (1-448) x US-09-815-242-3936 (1-1368)

```
Qy 11 lLeValValGlyLeuGlyLysSerGlyMetSerLeuValArgTyrLeuAlaArgGly 30
Db 37 TTGGTTTAGCATAGCTAAAGTGGTTCAGCGCAGCAACCTGATCAAAACCTGACGCCAAGATT 96
Qy 31 LeuProPheAlaValAlaAspThrArg-----GluAsnPro-----ProGluLeu 45
Db 97 GCGCTCGTTACCGTTAATGACGCAAAACAATTTGATCAAAACCTGACGCCAAGATT 156
Qy 46 AlaThrLeuArgAlaGlnTyrProGlnValGluValArgCysGlyGluLeuAspAlaGlu 65
Db 157 TTAACCTTG-----GGTATTCTGTTGTTTACAGGGGGGCGATCCCAATTGAA 201
Qy 66 PheLeuCysSerAlaArgGluLeuTyrVal---SerProGlyLeuSerLeuArgThrPro 84
Db 202 TTGTTGGATGAGAAATTTGAATCTGTTAAATAATCTGGTATTCCTTATCAAAATCCA 261
Qy 85 AlaLeuValGlnAlaAlaLysGlyValArgIleSerGlyAspIleAspLeuPheAla 104
Db 262 CTTTGGCAGAAAGCAGCTAACTCGGAAAATTCCTATCATCACTAGGTGGAATTTAGCAGGT 321
Qy 105 ArgGluAlaLysAlaProIleValAlaIleThrGlySerAsnAlaLysSerThrValThr 124
Db 322 CAAATTCGCAATGTCCAATTTGCGCATTCAGGGCACCATAATGGCAAAACCAACCGACCC 381
Qy 125 ThrLeuValGlyGluMetAlaValAlaAlaAspLys-----ArgValAlaValGly 141
Db 382 ACAGTATGTTGTT---TTACTGCTAAACGCTGACAGACGCGTGGTGGGACGCTTTGGCG 438
Qy 142 GlyAsnLeuGlyThrProAlaLeuAspLeu-----LeuAlaAspIle 156
Db 439 GGAAATATTGTTTTCAGCGAGTACGCTGGCTCAAGAAGCAACGCGCAGGATGATCTT 498
Qy 157 GluLeuTyrValLeuGluLeuSerSerPheGlnLeuGlnThrCysAspArgLeuAsnAla 176
Db 157 GluLeuTyrValLeuGluLeuSerSerPheGlnLeuGlnThrCysAspArgLeuAsnAla 176
```

```

Db 499 -----GTGATGGAACCTTCTAGCTTTTCACTTAATGGGAATGAGACGTTTCACCCA 549
QY 177 GluValAlaThrValLeuAsnValSerGluAspHisMetAspArgTyrAspGlyMetAla 196
Db 550 CAAATGTCAGTAATACAAATATTTTGAACACACTTGGATTATCATGTTCCGGGAAA 609
QY 197 AspTyrHisLeuAlaLysHisArgIlePheArgGlyAlaArg-----GlnValVal 213
Db 610 GAATATGTTGCTGCAAAATGGCCATTTCAAAAACATCATCGCTAGACACCTTGATT 669
QY 214 ValAsn-----ArgAlaAspAlaLeuThrArgProLeuIleAlaAspThrVal 229
Db 670 TTAATGCAATCAAGTAGAGCTTCAACAGTTAGCCAAACACACAGCTCCCAACGTATG 729
QY 230 ProCysTyrSerPheGlyLeuAsnLysProAspPheLysAlaPheGlyLeuIleGluGlu 249
Db 730 CCT-----TTTTCACGAAGAAGCAGTAGAAGGGCTTATCTTTTA----- 771
QY 250 AspGlyGlnTyrTrpLeuAlaPheGlnPheAspLysLeuLeuProValGlyGluLeuLys 269
Db 772 GATGGGAAA-----TTATATTTTATGAAGAATATATATATCCCGCATGAGCTAGG 825
QY 270 IleArgGlyValAlaHisAsnTyrSerAsnAlaLeuAlaAlaLeuAlaLeuGlyHisAlaVal 289
Db 826 ATCTCTGTAGTACAAATATGAAATGCACCTCGCAGCGATTGTGTAGCTAAATTAATA 885
QY 290 GlyLeuProPheAspAlaMetLeuGlyAlaLeuLysAlaPheSerGlyLeuAlaHisArg 309
Db 886 AATGTATCAATGTTTCAGATTAGACAAACTTTGAAAACTTTTCGGGCGTTCCTCATCGA 945
QY 310 CysGlnTyrValArgGlnGlyValSerTyrAspAspSerLysAlaThrAsn 329
Db 946 AGCAATTTGTTGGCGAAGTTCAGCAAGACGTTTTTATACGATTCAAACCAACCAAT 1005
QY 330 ValGlyAlaAlaLeuAlaAlaIleGluGlyLeuGlyAlaAspIleAspGlyLysLeuVal 349
Db 1006 ATTTAGCTACAGAGTGCCTTAAGTGGTTGAC-----AACCAAGACTACTT 1056
QY 350 LeuLeuAlaGlyGlyAspGlyLysGlyAlaAspPheHisAspLeuArgGluProValAla 369
Db 1057 TTACTTGGCGGTGGCTTGACCGCGGTACTCATTTGATGAATTG---GTTCTGCTGTG 1113
QY 370 ArgPheCysArgAlaValValLeuLeuGlyArgAspAlaGlyLeuIleAlaGlnAlaLeu 389
Db 1114 CTGGGACTCAAGCAATTTGTTTATTTGGAGAAACCAAGAAATATGCGGGAAGCTGCT 1173
QY 390 -----GlyAsnAlaValProLeuValArgValAlaThrLeuAspGluAlaValArgGln 407
Db 1174 AAAAAAGCGAACATTTGAACAATTTATTTGCTGAAATGTTCAACGCGGTTACCATT 1233
QY 408 AlaAlaGluLeuAlaArgGluGlyAspAlaValLeuLeuSerProAlaCysAlaSerLeu 427
Db 1234 GCCTTTGATTATTCGAAAAAGATGATACATATTTACTATCATCCTGCTTGCAGAGTTGG 1293
QY 428 AspMetPheLysAsnPheGluGluArgGlyArgLeuPheAlaLysAlaValGluGluLeu 447
Db 1294 GACCAATACCGAATTTTGAAGTACGCGGGGAAGCCTTTATGCAAGCTGTTCAACAATTA 1353

```

RESULT 15

US-09-815-242-6539  
 ; Sequence 6539, Application US/09815242  
 ; Patent No. US20020061569A1

GENERAL INFORMATION:

; APPLICANT: Haselbeck, Robert  
 ; APPLICANT: Ohlson, Kari L.  
 ; APPLICANT: Zyskind, Judith W.  
 ; APPLICANT: Wall, Daniel  
 ; APPLICANT: Trawick, John D.  
 ; APPLICANT: Carr, Grant J.  
 ; APPLICANT: Yamamoto, Robert T.  
 ; APPLICANT: Xu, H. Howard  
 ; TITLE OF INVENTION: Identification of Essential Genes in  
 ; TITLE OF INVENTION: Prokaryotes

```

; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-03-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6539
; LENGTH: 1371
; TYPE: DNA
; ORGANISM: Enterococcus faecalis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1371)
; US-09-815-242-6539

```

```

Alignment Scores:
Pred. No.: 1,368-51 Length: 1371
Score: 519.00 Matches: 152
Percent Similarity: 51.09% Conservative: 83
Best Local Similarity: 33.04% Mismatches: 181
Query Match: 22.96% Indels: 44
DB: 9 Gaps: 16

```

US-09-701-229-2 (1-448) x US-09-815-242-6539 (1-1371)

```

QY 11 lleValValGlyLeuGlyLysSerGlyMetSerLeuValArgTyrLeuAlaArgGly 30
Db 37 TTGGTTTAGGACTAGCTAAAGTGGTGTACGCGCAGCGAAACTCTTACATCAGTTAGT 96
QY 31 LeuProPheAlaValAlaAspThrArg-----GluAsnPro-----ProGluLeu 45
Db 97 GCGCTCGTTACGTTAATGACGCAAAACAATTTGATCAAAACCCCTGACGCCCAAGATT 156
QY 46 AlaThrLeuArgAlaGlnTyrProGlnValGluValArgCysGlyGluLeuAlaGlu 65
Db 157 TTAACCTTG-----GGTATTCGTGTTGTACAGGGGGGCATCCCAATTGAA 201
QY 66 PheLeuCysSerAlaArgGluLeuTyrVal----SerProGlyLeuSerLeuArgThrPro 84
Db 202 TTGTTGGATGAAGAATTTGAACATAATCGTTAAAAATCTGGTATCTCTATACAAATCCA 261
QY 85 AlaLeuValGlnAlaAlaLysGlyValArgIleSerGlyAspIleAspLeuPheAla 104
Db 262 CTTGTGGCAGAGCACTAACTCGGAAATTCCTATCAATCACTGAGGTGGAATTAGCAGGT 321
QY 105 ArgGluAlaLysAlaProIleValAlaIleThrGlySerAsnAlaLysSerThrValThr 124
Db 322 CAATTTGCCGAATGTCCCAATTTGCGGCATTTACGGCCACCAATGCGCAAAACAACACGACC 381
QY 125 ThrLeuValGlyGluMetAlaValAlaAlaAspLys-----ArgValAlaValGly 141
Db 382 ACGATGATTGTT--TTACTGCTAAACCTGACAGACGCGTGGTGGAGCAGCGTTTGGCG 438
QY 142 GlyAsnLeuGlyThrProAlaLeuAspLeu-----LeuAlaAspIle 156
Db 439 GGAATATTGTTTTCAGCGAGTACGCTGGTGGCTCAAGAAGCAACGCGCAAGGATGATCTT 498
QY 157 GluLeuTyrValLeuGluLeuSerPheGlnLeuGluThrCysAspArgLeuAla 176
Db 499 -----GTGATGGAACCTTCTAGTTTTCAGTTAATGGGAATTTGAGACGTTTCACCCA 549

```



Search completed: August 14, 2003, 11:04:37  
Job time : 6736 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus.p2n model

Run on: August 14, 2003, 07:04:21 ; Search time 2292 Seconds  
(without alignments)  
4750.613 Million cell updates/sec

Title: US-09-701-229-2  
Perfect score: 2260  
Sequence: 1 MSIIASHDFRIVVGLKSGM.....MFKNFEBGRFLFAKAVEELA 448

Scoring table:  
BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 22781392 seqs, 1215238056 residues

Total number of hits satisfying chosen parameters: 45546872

Minimum DB seq length: 25  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL=frame+ p2n.model -DEV=xlp  
-Q=/cgn2.1/USPTO.spool/US09701229/runat\_09082003.161157\_4393/app.query.fasta.1.647  
-DB=EST -QFWT=fastap -SUFFIX=p2n.rst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOCALIGN=200 -THR\_SCORE=pcpt -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=25 -MAXLEN=2000000000  
-USER=US09701229.ECGN.1.1.4237.runat\_09082003.161157\_4393 -NCPU=6 -ICPU=3  
-NO.MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV.TIMEOUT=120 -WARN.TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

- EST: \*
- 1: em\_estba:\*
  - 2: em\_esthum:\*
  - 3: em\_estin:\*
  - 4: em\_estnu:\*
  - 5: em\_estov:\*
  - 6: em\_estpl:\*
  - 7: em\_estro:\*
  - 8: em\_htc:\*
  - 9: gb\_est1:\*
  - 10: gb\_est2:\*
  - 11: gb\_htc:\*
  - 12: gb\_est3:\*
  - 13: gb\_est4:\*
  - 14: gb\_est5:\*
  - 15: em\_estfun:\*
  - 16: em\_estom:\*
  - 17: em\_gss\_hum:\*
  - 18: em\_gss\_inv:\*
  - 19: em\_gss\_pln:\*
  - 20: em\_gss\_vrt:\*
  - 21: em\_gss\_fun:\*
  - 22: em\_gss\_mam:\*
  - 23: em\_gss\_mus:\*
  - 24: em\_gss\_pro:\*
  - 25: em\_gss\_rod:\*
  - 26: em\_gss\_phg:\*
  - 27: em\_gss\_vri:\*
  - 28: gb\_gss1:\*

29: gb\_gss2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
c 1	1084	48.0	892	29	BZ579116	BZ579116 msh2_6205
c 2	1081.5	47.9	1445	29	BZ568959	BZ568959 pacs2-164
c 3	603.5	26.7	1171	29	BZ559806	BZ559806 pacs2-164
c 4	391	17.3	452	10	BE184674	BE184674 PM4-HT068
c 5	313	13.8	928	29	BZ567557	BZ567557 pacs2-164
c 6	240	10.6	714	28	BH374901	BH374901 AG-ND-127
c 7	233	10.3	748	12	BJ172647	BJ172647 BJ172647
c 8	231	10.2	760	28	BH367641	BH367641 AG-ND-127
c 9	226	10.0	1194	29	BZ550283	BZ550283 pacs1-60
c 10	213.5	9.4	572	28	BH795426	BH795426 BMBAC376G
c 11	210	9.3	965	29	CNS070MR	AL423817 T3 end of
c 12	187	8.3	595	29	CNS070MQ	AL423816 T7 end of
c 13	186	8.2	616	28	BH383040	BH383040 AG-ND-127
c 14	172.5	7.6	495	28	BH369957	BH369957 AG-ND-138
c 15	172	7.6	378	28	BH382726	BH382726 AG-ND-127
c 16	160.5	7.1	651	13	BQ504834	BQ504834 EST612249
c 17	136.5	6.0	689	10	BG592398	BG592398 EST491076
c 18	136	6.0	684	28	AQ968325	AQ968325 LFRJB56TR
c 19	132.5	5.9	1214	29	BZ572325	BZ572325 msh2_2585
c 20	129	5.7	422	29	BZ406343	BZ406343 OGAAAC90TF
c 21	125.5	5.6	838	29	CC115454	CC115454 NDL_42F3
c 22	124.5	5.5	794	14	CB853399	CB853399 UT-CF-FNO
c 23	124.5	5.5	930	29	BZ675115	BZ675115 PUBDE10TD
c 24	123	5.4	730	29	CC063642	CC063642 fgma003d0
c 25	122.5	5.4	146	9	AL719411	AL719411 AL719411
c 26	118	5.2	699	29	CC063650	CC063650 fgma003d0
c 27	117.5	5.2	685	29	CC140248	CC140248 NDL_42F4
c 28	114.5	5.1	191	14	CA548165	CA548165 CO802A11
c 29	112.5	5.0	817	29	CNS0110M	AL146223 Anopheles
c 30	110	4.9	1049	10	BE216920	BE216920 EST0463 T
c 31	106.5	4.7	584	10	BE760975	BE760975 SNQvL2CAS
c 32	104.5	4.6	673	12	BE718971	BE718971 1031033F0
c 33	103.5	4.6	884	29	BZ564204	BZ564204 pacs2-164
c 34	103.5	4.6	1108	29	BZ559252	BZ559252 pacs2-164
c 35	103.5	4.6	1482	29	BZ557157	BZ557157 pacs1-60
c 36	102.5	4.5	1522	12	BM320865	BM320865 rockefell
c 37	102	4.5	850	14	CB650740	CB650740 OSINEB15F
c 38	101.5	4.5	597	9	AW200691	AW200691 se92c10.y
c 39	101.5	4.5	876	10	BE216958	BE216958 EST0501 T
c 40	101.5	4.5	1031	29	BZ563054	BZ563054 pacs2-164
c 41	101	4.5	761	12	BM076245	BM076245 TrEST-A11
c 42	101	4.5	775	10	BG368290	BG368290 HVSME1001
c 43	101	4.5	829	10	BF264916	BF264916 HV_Cra001
c 44	101	4.5	1118	29	BZ548884	BZ548884 pacs1-60
c 45	100	4.4	822	11	CNS09MEH	EX065109 Single re

ALIGNMENTS

RESULT 1  
BZ579116/c  
LOCUS  
DEFINITION msh2\_6205, y2 msh Pseudomonas aeruginosa genomic clone msh2\_6205,  
892 bp DNA linear GSS 17-DEC-2002  
genomic survey sequence.  
ACCESSION BZ579116  
VERSION BZ579116.1 GI:27214177  
KEYWORDS GSS.  
SOURCE Pseudomonas aeruginosa  
ORGANISM Pseudomonas aeruginosa  
Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;  
Pseudomonadaceae; Pseudomonas.  
REFERENCE 1 (bases 1 to 892)

AUTHORS Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M.,  
Burns,J.L., Kaul,R. and Olsen,M.V.  
TITLE Whole-Genome-Sequence variation among multiple isolates of  
Pseudomonas aeruginosa library  
JOURNAL J. Bacteriol., (2002) In press  
COMMENT Contact: Chris K. Raymond  
Genome Center  
University of Washington  
Box 352145, Seattle, WA 98105-2145, USA  
Tel: 2062216954  
Fax: 2066857244  
Email: craymond@u.washington.edu  
Class: shotgun.

## FEATURES

Location/Qualifiers  
1..892

/organism="Pseudomonas aeruginosa"

/mol\_type="genomic DNA"

/strain="MSH"

/db\_xref="taxon:287"

/clone="msh2\_6205"

/clone\_lib="msh"

/note="Environmental isolate. Whole genomic shotgun  
library."

BASE COUNT 173 a 277 c 276 g 160 t 6 others  
ORIGIN

## Alignment Scores:

Pred. No.: 3,73e-112 Length: 892

Score: 1084.00 Matches: 230

Percent Similarity: 91.37% Conservative: 3

Best Local Similarity: 90.20% Mismatches: 19

Query Match: 47.96% Indels: 5

DB: 29 Gaps: 2

US-09-701-229-2 (1-448) x BZ579116 (1-892)

QY 23 ValArgTyrLeuAlaArgArgGlyLeuProPheAlaValAspThrArgGluAsnPro 42  
DB 839 GTGCGTTACTGG---CGCGGCGGNTGCTTTGGCGGTGCAATACC---GAAAGGANC 786  
QY 43 ProGlu-LeuAlaThrLeuAlaGlnTyrProGlnValGluValArgCysGlyGluLe 62  
DB 785 CCGGAAAGTGGCCACCTTGGTCCCGCATGTCG---CAAGTTGAAAGTGGTGGCGGCAACT 727  
QY 62 uAspAlaGluPheLeuCysSerAlaArgGluLeuTyrValSerProGlyLeuSerLeuAr 82  
DB 726 CAACGCGAGTTTCTTTTGTCTGCGCGCGAACTTTACGTGACG---CCCGGCTTGTCTGCG 668  
QY 82 gThrProAlaValGlnAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 102  
DB 667 CACCCCGCGCTGGTACAGCCGCCGCGGAAAGCGCGCATCTCCGTTGACATCGATCT 608  
QY 102 uPheAlaArgGluAlaLysAlaProIleValAlaIleThrGlySerAsnAlaLysSerTh 122  
DB 607 CTTCCGCGCGAGGCGAGGCGAGGCGGATCGTCCCATCACCGTTCCACGCGGAGAGCAC 548  
QY 122 rValThrThrLeuValGlyGluMetAlaValAlaAlaAlaAlaAlaAlaAlaAlaAla 142  
DB 547 CGTGACCACTCTGTGTGGGAAATGGCGTGGCGCGGACAAAGCGTGTGCGCGTGGCGGG 488  
QY 142 yAsnLeuGlyThrProAlaLeuAspLeuLeuAlaAlaAlaAlaAlaAlaAlaAlaAla 162  
DB 487 CAACTCGGCGCGCGCGCTGACCTGTGGCGCGGACGACATCGAGCTGTACGTTGGGA 428  
QY 162 uLeuSerSerPheGlnLeuGluThrCysAspArgLeuAsnAlaGluValAlaIleThrVal 182  
DB 427 GCTGTGAGCTTCCAGCTGGAACCTGCGATCGCTCAACGCGGAGGTGGCGACCTGCT 368  
QY 182 uAsnValSerGluAspHisMetAspArgTyrAspGlyMetAlaAspTyrHisLeuAlaLy 202  
DB 367 GAACGTCGCGAAGACCATATGATGATGCTAGCAGCGCATGGTGTACTACCACTGGCCAA 308  
QY 202 sHisArgIlePheArgGlyAlaArgGlnValValValAsnArgAlaAspAlaLeuThrAr 222

DB 307 GCACCGGATCTTCCGCGGTGCCCGCAGTCTGGTGAATCGCGCGATGCTGACCGG 248  
QY 222 gProLeuIleAlaAspThrValProCysTrpSerPheGlyLeuAsnLysProAspPheLy 242  
DB 247 ACCGCTGATCGCGCATACCGTCCGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 188  
QY 242 sAlaPheGlyLeuIleGluAlaAspGlyGlnLysTrpLeuAlaPheGlnPheAspLysLe 262  
DB 187 GCGTTTGGCGCTGATCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 128  
QY 262 uLeuProValGlyGluLeuLysIleArgGlyAlaAlaHisAsnTyr 276  
DB 127 GCTGCCGCTTGGCGAACTGAAGATCCGTGGCGGCCCAACTAT 85

## RESULT 2

BZ568959/c

LOCUS

DEFINITION

pac2-164\_8171.y2 pac2-164 Pseudomonas aeruginosa genomic clone

pac2-164\_8171, genomic survey sequence.

ACCESSION BZ568959

VERSION BZ568959.1 GI:27202799

KEYWORDS GSS.

SOURCE Pseudomonas aeruginosa

ORGANISM Pseudomonas aeruginosa

REFERENCE 1 (bases 1 to 1445)

AUTHORS Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M.,

Burns,J.L., Kaul,R. and Olsen,M.V.

TITLE Whole-Genome-Sequence variation among multiple isolates of

Pseudomonas aeruginosa library

J. Bacteriol., (2002) In press

COMMENT Contact: Chris K. Raymond

Genome Center

University of Washington

Box 352145, Seattle, WA 98105-2145, USA

Tel: 2062216954

Fax: 2066857244

Email: craymond@u.washington.edu

Class: shotgun.

FEATURES

Location/Qualifiers

1..1445

/organism="Pseudomonas aeruginosa"

/mol\_type="genomic DNA"

/strain="2-164"

/db\_xref="taxon:287"

/clone="pac2-164\_8171"

/clone\_lib="pac2-164"

/note="clinical isolate 2-164 Whole genomic shotgun

library."

BASE COUNT 308 a 424 c 412 g 293 t 8 others

ORIGIN

Alignment Scores:

Pred. No.: 1.49e-111 Length: 1445

Score: 1081.50 Matches: 241

Percent Similarity: 87.63% Conservative: 7

Best Local Similarity: 85.16% Mismatches: 25

Query Match: 47.85% Indels: 15

DB: 29 Gaps: 4

US-09-701-229-2 (1-448) x BZ568959 (1-1445)

QY 69 SerAlaArgGluLeuTyrVal-----SerProGlyLeuSerLeuArgThr 83

DB 904 AACTCACCAGAGTCTTTTGTGGCGGCAACTTTACTACCGCGGCTTCTGTTGTC-TGC 846

QY 84 ProAlaLeuValGlnAlaAlaLysGlyValArgIleSerGlyAspIleAspLeuPhe 103

DB 845 CCGCGCTGTACAGCGCGCGGAAAGCGTGGCATCTCGTAC-----ATCGATCTT--- 798

QY 104 AlaArgGluAlaLysAlaProIleValAlaIleThrGlySerAsnAlaLysSerThrVal 123

```

Db 797 TCAGCGCAGGCGGCGCGA---TCGTNGCATCACGGTTCCACGCGGAAGACCGCTG 741
Qy 124 ThrThrLeuValGlyGluMetAlaValAlaAspLysArgValAlaValAlaGlyGlyAsn 143
Db 740 ACAACCTT-GTGGCGGAATGACGGTGGC-GCGNACAAAGCGTGTGCGCTC-GNCGGCAAC 684
Qy 144 LeuGlyThrProAlaLeuAspLeuLeuAlaAspAspIleGluLeuThrValLeuGluLeu 163
Db 683 CTCGGCAC-CCGGCGCTCGACCTGCTGCCGACACATCGAGCTGTACGTGTGGAGCTG 625
Qy 164 SerSerPheGlnLeuGluThrCysAspArgLeuAsnAlaGluValAlaThrValLeuAsn 183
Db 624 TCAGAGCTTCAGCTGGAACCTCGCATCGCTCAACGCCGAGGTGGCGACCGTCTGAAC 565
Qy 184 ValSerGluAspHisMetAspArgTyrAspGlyMetAlaAspTyrHisLeuAlaLysHis 203
Db 564 GTACGCGAAGACCATATGGATCGCTACGCGCATGGCTGACTACCACTGGCGCAACGAC 505
Qy 204 ArgIlePheArgGlyAlaArgGlnValValValAsnArgAlaAspAlaLeuThrArgPro 223
Db 504 CGGATCTTCGCGGTGCCGCGCAGGTGCTGTAATCGCGCGATGCGCTGACCGACCG 445
Qy 224 LeuIleAlaAspThrValProCysTrpSerPheGlyLeuAsnLysProAspPheLysAla 243
Db 444 CTGATCGCGGATACCGTGGCGGTGCTGCTGCGCTGAACAGCGGACTTCAAGGCT 385
Qy 244 PheGlyLeuIleGluAspGlyGlnLysTrpLeuAlaPheGlnPheAspLysLeuLeu 263
Db 384 TTCGGCTGTATCGAGGAGAGCGGCAGAGTGGTGGCTTCAGTTCGACAGCTGCTG 325
Qy 264 ProValGlyGluLeuLysIleArgGlyAlaHisAsnTyrSerAsnAlaLeuAlaAlaLeu 283
Db 324 CCGGTTCGGCAACTGAAGATCCGTGGCGGCCACAACTATTCCAAACGCGCTCGCGCGCTG 265
Qy 284 AlaLeuGlyHisAlaValGlyLeuProPheAspAlaMetLeuGlyAlaLeuLysAlaPhe 303
Db 264 GCGTGGGCGCATCGGTGGCGCTGCCGTTCGACGCCATGCTCGCGCGCTGAAGCGCGTTT 205
Qy 304 SerGlyLeuAlaHisArgCysGlnTrpValArgGluArgGlnGlyValSerTyrTyrAsp 323
Db 204 TCCGGCTGTGCTATCGCTGCCAGTGGTACGCGAGCGGCGGCGGTGAGCTACTAGCAC 145
Qy 324 AspSerLysAlaThrAsnValGlyAlaAlaLeuAlaAlaIleGluGlyLeuGlyAlaAsp 343
Db 144 GATTCCAGGCCACCAACGTCGGCGCGCGCTGCGCGCGATCGAGGGGCTGGGTGCGGAC 85
Qy 344 IleAspGly 346
Db 84 ATCGACGGG 76

```

## RESULT 3

BZ559806 1171 bp DNA linear GSS 17-DEC-2002  
 LOCUS pacs2-164\_1863.x1 pacs2-164 Pseudomonas aeruginosa genomic clone  
 DEFINITION pacs2-164\_1863, genomic survey sequence.

## ACCESSION

BZ559806

## VERSION

BZ559806.1 GI:27176698

## KEYWORDS

GSS.

## SOURCE

Pseudomonas aeruginosa

## ORGANISM

Pseudomonas aeruginosa

## REFERENCE

1 (bases 1 to 1171)

## AUTHORS

Spencer, D.H., Raymond, C.K., Smith, E.E., Sims, E.E., Hastings, M.,

## TITLE

Burns, J.L., Kaul, R. and Olsen, M.V.

## JOURNAL

Whole-Genome-Sequence variation among multiple isolates of

## COMMENT

Pseudomonas aeruginosa library

## JOURNAL

J. Bacteriol., (2002) In press

## COMMENT

Contact: Chris K. Raymond

Tel: 2062216954

Fax: 2066857244

Email: craymond@u.washington.edu

Class: shotgun

## FEATURES

## source

1. 1171  
 Location/Qualifiers  
 /organism="Pseudomonas aeruginosa"  
 /mol\_type="genomic DNA"  
 /strain="2-164"  
 /db\_xref="taxon:287"  
 /clone="pacs2-164\_1863"  
 /clone\_lib="pacs2-164"  
 /note="clinical isolate 2-164 Whole genomic shotgun library."

BASE COUNT 218 a 325 c 329 g 298 t 1 others

## ORIGIN

Alignment Scores:  
 Pred. NO.: 1.64e-57 Length: 1171  
 Score: 603.50 Matches: 132  
 Percent Similarity: 83.13% Conservative: 6  
 Best Local Similarity: 79.52% Mismatches: 25  
 Query Match: 26.70% Indels: 3  
 DB: 29 Gaps: 1

US-09-701-229-2 (1-448) x BZ559806 (1-1171)

Qy 276 TyrSerAsnAlaLeuAlaAlaLeuAlaLeuGlyHisAlaValGlyLeuProPheAspAla 295  
 Db 48 TATTCCAAACGCGCTCGCGGAGCTGGACATGGGCGCATGGGTCGCGCTGTTACGCC 107  
 Qy 296 MetLeuGlyAlaLeuLysAlaPheSerGlyLeuAlaHisArgCysGlnTrpValArgGlu 315  
 Db 108 ATGCTCGCGCGCTGAAAGCGTCTTGGGACTGCTCATTTGCTCCAGTGGGTACGCGAG 167  
 Qy 316 ArgGlnGlyValSerTyrTyrAspSerLysAlaThrAsnValGlyAlaAlaLeuAla 335  
 Db 168 CGGCGAGGCGGTGAGCTACTAGCAGCATTTCCAAAGGCCACCAACGTCGCGCGCTGCGG 227  
 Qy 336 AlaIleGluGlyLeuGlyAlaAspIleAspGlyLysLeuValLeuLeuGlyGlyAsp 355  
 Db 228 GCGATCAGAGGGCTGGGTGGCGCATGTCAGCGGCAAGCTGTGTGTGTCGCGCGGAGAG 287  
 Qy 356 GlyLysGlyAlaAspPheHisAspLeuArgGluProValAlaArgPheCysArgAlaVal 375  
 Db 288 GGCAAGGCGCGGATTTTCATGACCTGCGGAGCGGTGCGCGCTTCTGCGCGCGGTG 347  
 Qy 376 ValLeuLeuGlyArgAspAlaGlyLeuIleAlaGlnAlaLeuGlyAsnAlaValProLeu 395  
 Db 348 GTACTGCTTGTGCTGACACCTAGCTGATTGACGAGCAGCTGGGCAACGCGGACCGCTG 407  
 Qy 396 ValArgValAlaThrLeuAspGluAlaValArgGlnAlaAlaGluLeuAlaArgGluGly 415  
 Db 408 GTGCGCGTCAACGCTGGAGAAAGCATTTCCGAGGCGCGCGAGCTGGCGCGCGGAAGC 467  
 Qy 416 Asp-AlaValLeuLeuSerProAlaCysAla---SerLeuAspMetPhe-LysAsnPheG 434  
 Db 468 GATAATAGTCTCTGGGATTCCCGGCCCTTCCCAACCCGCTGATATGTATAAANAACCTTT 527  
 Qy 434 luGluArgGlyArg 438  
 Db 528 TAAAAAAGGCGG 541

## RESULT 4

## LOCUS

BE184674

## DEFINITION

PM4-HT0688-050500-002-d10 HT0688 Homo sapiens cDNA, mRNA sequence.

## ACCESSION

BE184674

## VERSION

BE184674.1 GI:8663858

## KEYWORDS

EST.

## SOURCE

Homo sapiens (human)

## ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

452 bp mRNA linear EST 22-JUN-2000

BE184674

PM4-HT0688-050500-002-d10 HT0688 Homo sapiens cDNA, mRNA sequence.

BE184674

BE184674.1 GI:8663858

EST.

Homo sapiens (human)

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Db	262	GC	ACTCGGTACCGCGTTGGCTGCCATTCGACGCCATGCTCGCAGACCTGCCACGAT	321
Qy	304	SerGlyLeuAlaHisArgCysGlnTtpValArgGluArgInGlyValSerTyr-TyrAs	323	
Db	322	CGCGGCTTGAGCATCGCTGCCAATGGTGGCGGATCTCAATCAAGTCACGCTATTACAA	381	
Qy	323	AspSerLyAlaThraSsnValGlyAla-AlaLeuAlaAlaIleGluGlyLeuGlyAlaA	343	
Db	382	TGACTCAAAGCCACCAACGTCGTCGGCCCTTGAGGCCATCGCAGGCTGGTGGCGC	441	
Qy	343	spile 344		
Db	442	ATATC 446		
RESULT 5				
BZ567557/C				
LOCUS	BZ567557	928 bp	DNA	linear
DEFINITION	pac2-164_7094.y2 pac2-164 Pseudomonas aeruginosa genomic clone			GSS 17-DEC-2002
ACCESSION	pac2-164_7094			genomic survey sequence.
VERSION	BZ567557			
KEYWORDS	BZ567557.1	GI:27199005		
SOURCE	GSS.			
ORGANISM	Pseudomonas aeruginosa			
REFERENCE	Pseudomonas aeruginosa			
AUTHORS	Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales; Pseudomonadaceae; Pseudomonas.			
TITLE	1 (bases 1 to 928)			
JOURNAL	Spencer, D.H., Raymond, C.K., Smith, E.E., Sims, E.E., Hastings, M., Burns, J.L., Kaul, R. and Olsen, M.V.			
COMMENT	Whole-Genome-Sequence variation among multiple isolates of Pseudomonas aeruginosa library J. Bacteriol., (2002) In press Contact: Chris K. Raymond Genome Center University of Washington Box 352145, Seattle, WA 98105-2145, USA Tel: 2062216954 Fax: 2066857244 Email: craymondeu.washington.edu Class: shotgun.			
FEATURES				
source	Location/Qualifiers			
	1..928			
	/organism="Pseudomonas aeruginosa"			
	/mol_type="genomic DNA"			
	/strain="2-164"			
	/db_xref="taxon:287"			
	/clone="pac2-164_7094"			
	/clone_lib="pac2-164"			
	/note="clinical isolate 2-164 Whole genomic shotgun library."			
BASE COUNT	238 a	228 c	236 g	226 t
ORIGIN				
Alignment Scores:				
Pred. No.:	9.72e-25	Length:	928	
Score:	313.00	Matches:	89	
Percent Similarity:	60.1%	Conservative:	24	
Best Local Similarity:	47.34%	Mismatches:	65	
Query Match:	13.85%	Indels:	10	
DB:	29	Gaps:	3	
US-09-701-229-2 (1-448) x BZ567557 (1-928)				
Qy	193	AspGlyMetAlaHisArgCysGlnTtpValArgGluArgInGlyValSerTyr-TyrAs	212	
Db	581	GATGGGTG---AATATCTTTGGCCCAACATCGGTTCTTAATACGTTGGCGTCAGGT	525	
Qy	212	IValValAsn-ArgAlaAspAlaLeuThrArgProLeuIleAlaAspThrValProCysT	232	
Db	524	CTGGTTAATTCGGCGTGATGCTCGACCCGCGCTGATCGTCACGTCACCGTACCGTACT	465	
Qy	232	rpSerPheGlyLeuAsnLyProAspPheLysAlaPheGlyLeuIleGluGlyAspGlyC	252	

```

Db 464 GGTGTTGACCTCAGCAGCGGCTTTACAAAGCTTTTCGACCTGGTCCAGAAAGTCGTAC 405
Qy 252 lnLysTrpLeuAlaPheGlnPheAspLysLeuLeuProValGlyGluLeuLysIleArg 272
Db 404 AGAAGTGGCTGGCTAGCTGTTGACGAAATAATTCGGCTGCTGATGTGAAGATCGTC 345
Qy 272 yAlaAlaHisAsnTyrSerAsnAlaLeuAlaLeuAlaLeuGlyHisAlaValGlyLeu 292
Db 344 GAGCAGCGGCTATTCATACGCTTCCGCGCTGGGCTTGGGCGCATCGCTCGTCGTG 285
Qy 292 rPhe-AspAlaMetLeuGlyAlaLeuLysAlaPheSerGlyLeuAlaHisArgCysGln 311
Db 284 GGTTCACGACCATATGATGCGGCTACAGAGTGCACGCTGATCCAGCTTCTCTATCCTCGCCA 225
Qy 312 TrpValArgGluArgGlnGlyValSerTyrTyr-AspAspSerLysAlaThrAsnValG1 331
Db 224 TGAGTACCCCTAACGACAGGCTCTGATGATGTCGGACGATTCACGAGCCGCTCTGGTCTG 165
Qy 331 yAlaAlaLeuAlaAlaIleGluGlyLeuGlyAlaAspIleAspGlyLysLeuValLeu 351
Db 164 CACGCGCCTGTTTCGATCGAGTGGATGGTGTGGACATTTTCATCAACCTCATCTGTG 105
Qy 351 uAlaGlyClyAspGlyLysGlyAlaAspPheHisAspLeuArgGluProValAlaArg-- 370
Db 104 CGCAGACGACGACGACGAGGGG---GATCCACTAGTTCTAGAGCGCGCCGACCGCGTG 48
Qy 371 -----PheCys 372
Db 47 GAGCTCCAGCTTTTGT 32

RESULT 6
BH374901/c
LOCUS
DEFINITION AG-ND-127J23-TR ND-TAM Anopheles gambiae genomic clone AG-ND-127J23
ACCESSION BH374901
VERSION BH374901.1 GI:17321043
KEYWORDS GSS.
SOURCE Anopheles gambiae (African malaria mosquito)
ORGANISM Anopheles gambiae
REFERENCE 1 (bases 1 to 714)
AUTHORS Hong, Y.S., Hogan, J.R., Wang, X., Sarkar, A., Sim, C., Loftus, B.J., Ren, C., Huff, E.R., Carlile, J.L., Black, K., Zhang, H.-B., Gardner, M.J. and Collins, F.H.
TITLE Construction of a BAC library and generation of BAC end
JOURNAL malaria mosquito Anopheles gambiae
MEDLINE Mol. Genet. Genomics 268 (6), 720-728 (2003)
PUBMED 22542063
COMMENT Other GSSs: AG-ND-127J23.TF
Contact: Brendan J Loftus
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0208
Fax: 301 838 3543
Email: b.loftus@tigr.org
This clone is from an A. gambiae BAC library (ND-TAM) provided by
F.H. Collins and sequenced by The Institute for Genomic Research
(TIGR). The BAC library was generated from A. gambiae PEST strain
DNA. All DNA was extracted from newly hatched first instar larvae
to minimize the inclusion of DNA from microorganisms that inhabit
the gut. The DNA is derived from mixed sexes of larvae. The BAC
library was constructed at Texas A&M University BAC Center
University, College Station, Texas 77843-2123, USA using a HindIII
partial digest.
Seq primer: M13 Rev
Class: BAC ends.
Location/Qualifiers

```

```

source
1. 714
/organism="Anopheles gambiae"
/mol_type="genomic DNA"
/strain="pEST"
/db_xref="taxon:7165"
/clone="AG-ND-127J23"
/clone_lib="ND-TAM"
/note="Vector: pECBAC1; Site_1: HindIII"
BASE COUNT 212 a 140 c 106 g 256 t
ORIGIN
Alignment Scores:
Pred. NO.: 1.23e-16 Length: 714
Score: 240.00 Matches: 67
Percent Similarity: 46.56% Conservative: 48
Best Local Similarity: 27.13% Mismatches: 94
Query Match: 10.62% Indels: 38
DB: 28 Gaps: 7
US-09-701-229-2 (1-448) x BH374901 (1-714)
Qy 171 CysAspArgLeu-----AsnAlaGluValAlaThrValLeuAsnValSerGluAsp 187
Db 706 TGTGATAATATTCAGTATTTTCAGACCATATATCAGTTGCTCTCTAAATCTGTCTCAGGAC 647
Qy 188 HisMetAspArgTyrAsp---GlyMetAlaAspTyrHisLeuAlaLysHisArgIlePhe 206
Db 646 CATTTAGACCAGTATTAATAATATGAAAGTACGATTAAGCAAAATTTCTGATCTTTCA 587
Qy 207 ArgGlyAlaArgGln-----ValValValAsnArgAlaAspAlaLeuThrArgPro 223
Db 586 GAAATCAGGAAAATGATAATTTATTCATCTACAATAAAGATGATGAATGAGCCAGAAA 527
Qy 224 LeuIle-----AlaAspThrValProCysTyrTrpSerPheGlyLeuAsn 237
Db 526 ATTCTTCAGTCTCTGGAGATTAAATGCTACAATGATTCCT----- 488
Qy 238 LysProAspPheLysAlaPheGlyLeuIleGlu----- 249
Db 487 -----TTCTCTATGAACAGAGAGCTAAGTGAAGGTGTTATTCAATT 446
Qy 250 AspGlyGlnLysTrpLeuAlaPheGlnPheAspLysLeuLeuProValGlyGluLeuLys 269
Db 445 GACAATGAATTAGTCATTAATACTTCAGGACGACTTCAGAATGAAATTTCTGATCTTTCA 386
Qy 270 IleArgGlyAlaHisAsnTyrSerAsnAlaLeuAlaLeuAlaLeuGlyHisAlaVal 289
Db 385 CTTGTAGAAACCATATAATGTTGCCAATAGTTTACGTCAGTATAGCAGGTAACACTGCTA 326
Qy 290 GlyLeuProPheAspAlaMetLeuGlyAlaLeuLysAlaPheSerGlyLeuAlaHisArg 309
Db 325 AATATTAGTAATGAAAGTATCCGCAATAGTCTGATGACATTTTCAGGCTGTCCGCACAGA 266
Qy 310 CysGlnTrpValArgGluArgGlnGlyValSerTyrTrpAspAspSerLysAlaThrAsn 329
Db 265 TTGGAGCAGGTGTCAGTTATTAATGATGTAATACATCAATGACAGTAAGGCAACCAAT 206
Qy 330 ValGlyAlaAlaLeuAlaAlaIleGluGlyLeuGlyAlaAspIleAspGlyLysLeuVal 349
Db 205 GTAATGCTGATATTTATGCACTGAA-----AGTGTGAATATCTCTGCAATA 158
Qy 350 LeuLeuAlaGlyLysGlyAspGlyLysAlaAspPheHisAspLeuArgGluProValAla 369
Db 157 TGGATTGTTGCTGGAATAGATAAGAAATCACTACAGAAATTTGAGGATCTGTGTAAG 98
Qy 370 ArgPheCysArgAlaValValLeuLeuGlyArgAspAlaGlyLeuIleAlaGlnAlaLeu 389
Db 97 AAGAAAGTTCGTGCGATCGTATGCTGGGATTGGATATGAAAGATCATTTCACTCTTC 38
Qy 390 GlyAsnAlaValProLeuVal 396
Db 37 AGAAACAAAAGGATTGTTGATT 17

```

FEATURES







```

Db      466 CAAGACTTCGAAGACCGCAACGCTGTCGCCAAACCGTAGAGGAGCTAGCG 412

RESULT 10
BH795426/c
LOCUS   572 bp      DNA      linear      GSS 09-APR-2002
DEFINITION   BMBAC376G09P7.PSU Brugia malayi Genomic Bac Library 3 Brugia malayi
              genomic; genomic survey sequence.
ACCESSION   BH795426
VERSION     BH795426.1 GI:20043758
SOURCE      GSS.
           Brugia malayi
           Brugia malayi
           Brugia malayi
           Eukaryota; Metazoa; Nematoda; Chromadorea; Spirurida; Filarioidea;
           Onchocercidae; Brugia.
REFERENCE   1 (bases 1 to 572)
AUTHORS     Whittom,C., Daub,J., Ware,J., Quail,M., Hall,N., Barrell,B., Foster
              J., Guiliano,D., Slatko,B. and Blaxter,M.
TITLE       Genome survey sequences from the human parasitic nematode Brugia
              malayi
JOURNAL     Unpublished
COMMENT     Contact: Blaxter ML
              Institute of Cell, Animal and Population Biology
              University of Edinburgh
              Ashworth Labs, King's Buildings, West Mains Road, Edinburgh, EH9
              3JT, UK
              Tel: +44 131 650 6760
              Fax: +44 131 670 5450
              Email: mark.blaxter@ed.ac.uk
              Sequenced from the Brugia malayi BAC library constructed by Claire
              Whittom and Dr Mike Quail. The sequence was generated by The
              Pathogen Sequencing Unit, The Sanger Institute, Cambridge, UK in
              collaboration with Mark Blaxter, ICAPB, University of Edinburgh,
              Edinburgh, UK
              Seq primer: T7 (TAATACGACTCACTATAGG)
              Class: BAC ends.

FEATURES             Location/Qualifiers
     source           1..572
                     /organism="Brugia malayi"
                     /mol_type="genomic DNA"
                     /strain="TRS"
                     /db_xref="taxon:6279"
                     /sex="Mixed (male and female)"
                     /tissue_type="whole parasite"
                     /dev_stage="microfilaria (L1)"
                     /clone_lib="Brugia malayi Genomic Bac Library 3"
                     /note="Vector: pBACe3.6; Site_1: BamH I; Brugia malayi
                     genomic DNA was partially cleaved with Sau3A I and size
                     fractionated. 7,392 clones were generated with mean insert
                     size ~48 kbp. The library was constructed by Claire
                     Whittom, Blaxter Nematode Genetics Lab, University of
                     Edinburgh, UK, and Dr Mike Quail, The Pathogen Sequencing
                     Unit, The Sanger Centre, Cambridge, UK."

BASE COUNT  154 a 120 c 86 g 212 t
ORIGIN

Alignment Scores:
Pred. No.:      8,84e-14      Length:      572
Score:          213.50      Matches:      56
Percent Similarity: 61.1%      Conservative: 21
Best Local Similarity: 44.4%      Mismatches:  31
Query Match:     9.45%      Indels:      18
DB:              28          Gaps:          6

US-09-701-229-2 (1-448) x BH795426 (1-572)

QY      52 TyrProGlnValGluValArgCysGly-GluLeuAspAlaGluPheLeuCysSerAlaAr 71
       ||||| : : : : : |||
Db      504 TATCCATCAAAAATACGATTGCGACGAGATAAAGCA----- 465

QY      71 gGluLeuTyValSerProGlyLeuSerLeuArgThrProAla-----LeuValGl 88
       ||| : : : : : ||| : : :
Db      464 ----TTGATTTTAAGCCCTGGAGTACCAATTTTCATCCGAGCCACATGGATAGTAA 409

```

```

QY      88 nAlaAlaAlaLys---GlyValArgIleSerGlyAspIleAspLeuPheAlaArgGluAl 107
       : ||| : : : : : ||| : : : : : ||| : : : : : |||
Db      408 ACTTGCAGAAGATTTTGGACTGTGTAATAAAATCGGATATTGAACTATTCTA---GAAG 352

QY      107 aLysAla-----ProIleValAlaIleThrGlySerAspAlaLysSerThrValTh 124
       ||||| : : : : : ||| : : : : : ||| : : : : : |||
Db      351 TAAAGCTAAGACCAACGAAAATTTGAGCGTCACAGGAACGAATGGCAAAATCAACCACTAC 292

QY      124 rThrLeuValGluMetAlaValAlaAlaAspLysArgValAlaValAlaValGlyGlyAsnLe 144
       : : : : : : : : : : : : : : : : : : : : : : : :
Db      291 GTCATTAAATAGCTACATATTAAATCTCGGGGAAAAAAGTAGCTATTGGTGAATTT 232

QY      144 uGlyThrProAlaLeuAspLeuAlaAspIleGluLeuTyValLeuGluLeuSe 164
       ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      231 AGGTGTTCTATTGTTGAT---CTAGAAAAAGATCGGAAATTTATGTAAATGAAATTC 175

QY      164 rSerPheGlnLeuGlu 169
       ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      174 CTCCTTTCAATTAGAA 159

RESULT 11
CNS070MR/c
LOCUS   965 bp      DNA      linear      GSS 06-JUL-2001
DEFINITION   T3 end of clone AZ0AA010F07 of library AZ0AA from strain CBS 712 of
              Kluyveromyces marxianus, genomic survey sequence.
ACCESSION   AL423817
VERSION     AL423817.1 GI:12207011
KEYWORDS    GSS.
SOURCE      Kluyveromyces marxianus
ORGANISM    Kluyveromyces marxianus
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; Saccharomycetaceae; Kluyveromyces.
REFERENCE   1 (bases 1 to 965)
AUTHORS     Souciet,J.L., Aigle,M., Artiguenave,F., Blandin,G.,
              Bolotin-Fukuhara,M., Bon,E., Brottier,P., Casaregola,S.,
              de-Montigny,J., Dujon,B., Durrien,P., Lepingle,A., Liorente,B.,
              Malpertuy,A., Neuveglise,C., Ozier-Kalogeropoulos,O., Potier,S.,
              Saurin,W., Tekala,F., Toffano-Nioche,C., Wesolowski-Louvel,M.,
              Wincker,P. and Weissenbach,J.
              Genomic exploration of the hemiascomycetous yeasts: 1. A set of
              yeast species for molecular evolution studies
              FEBS Lett. 487 (1), 3-12 (2000)
              20584711
              11152876
              2 (bases 1 to 965)
              and Dujon,B.
              Genomic exploration of the hemiascomycetous yeasts: 12.
              Kluyveromyces marxianus var. marxianus
              FEBS Lett. 487 (1), 71-75 (2000)
              20584722
              11152887
              3 (bases 1 to 965)
              Genoscope.
              Direct Submission
              Submitted (08-SEP-2000) Genoscope - Centre National de Sequencage,
              2 rue Gaston Cremieux, Cp 5706, 91057 EVRY cedex, FRANCE. (E-mail :
              seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
              This GSS is part of a random genomic sequencing program of thirteen
              yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces
              exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii,
              Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces
              lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia
              angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila,
              Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to
              5 kb were prepared and both extremities were sequenced. See
              keywords for description of this sequence and for the sequence of
              the other extremity of this insert.
              Location/Qualifiers
              1..965
              /organism="Kluyveromyces marxianus"
              /mol_type="genomic DNA"
              /strain="CBS 712"

FEATURES             source

```

```

/variety="marxianus"
/db_xref="taxon:4911"
/clone="AZ0AA010F07"
/clone_lib="AZ0AA"
/notes="end : T3"
complement(<355..>836)
/notes="similar to CAB08672 [ murD ]
1 putative frameshift(s)"
/evidence=not_experimental
misc_feature 199 a 275 c 310 g 159 t 22 others
ORIGIN

```

```

Alignment Scores:
Pred. No.: 4.88e-13 Length: 965
Score: 210.00 Matches: 59
Percent Similarity: 50.55% Conservative: 33
Best Local Similarity: 32.42% Mismatches: 77
Query Match: 9.29% Indels: 15
DB: 29 Gaps: 2

```

US-09-701-229-2 (1-448) x CNS070MR (1-965)

```

Qy 274 HisAsnTyrSerAsnAlaLeuAlaLeuAlaLeuGlyHisAlaValGlyLeuProphe 293
Db 883 CACASATGTCGAGACCGCTGGCGCAGCTGCGCTAACTCGCTCTTC-GGCGTCCGCC 825
Qy 294 AspAlaMetLeuGlyAlaLeuLysAlaPheSerGlyLeuAlaHisArgCysGlnTrpVal 313
Db 824 CGCGCCGTGGTCAAGGTCTCGCGGACCTCCACCTTSGCGGCATCGTATCGAGACGTC 765
Qy 314 ArgGluArgGlnGlyValSerTyrTyrAspSerLysAlaThrAsnValGlyAlaAla 333
Db 764 CATAAAGCGGTGGCATCACSKGGKCGAGACTCCAAAGGCCACCCWATYTTTTTTCGCG 705
Qy 334 LeuAlaAlaIleGluGlyLeuGlyAlaAspIleaspGlyLysLeuValLeuAlaGly 353
Db 704 TTCTCTCTWTGCGCGCTTC-----GAGCACATCTTGTGATCGCCGGA 660
Qy 354 GlyAspGlyLysGlyAlaAspPheHisAspLeuArgGluProValAlaArgPheCysArg 373
Db 659 GGCCAGGCCAAGGCGACCTACTTTGAGGACCTGTGTACCCACCGCCGAGAGTGCGC- 601
Qy 374 AlaValValLeuGlyArgAspAlaGlyLeuIleAlaGlnAlaLeuGlyAsnAlaVal 393
Db 600 GGAGCCATCGTGTGGAAACCGACCGCTGACGTGCTGCCAAGTCNTTGGCCGAGCAGCC 541
Qy 394 ProLeuValArgValAlaThrLeuAspGlu-----AlaVal 405
Db 540 CCGCATGTCCTCGCTCGTCATGATGCTTGTCCGAGATGCTATGCGCCAGGCGAGTT 481
Qy 406 ArgGlnAlaAlaGluLeuAlaArgGluGlyAspAlaValLeuLeuSerProAlaCysAla 425
Db 480 CGCGAGGCTGCGAGATGGCCCGCCGCGGACACCGCTTGTGATGCTCGCGGGTGGCG 421
Qy 426 SerLeuAspMetPheLysAsnPheGluArgGlyArgGlyLeuPheAlaLysAlaValGlu 445
Db 420 AGCCTCGAAATTTGGAATGATATGATGMMGCGGTGACGATTTCTGTCAGCGCGGACGA 361
Qy 446 GluLeu 447
Db 360 CAAATT 355

```

```

RESULT 12
CNS070MQ
LOCUS
DEFINITION T7 end of clone AZ0AA010F07 of library AZ0AA from strain CBS 712 of
Kluyveromyces marxianus, genomic survey sequence.
ACCESSION
AL423816
VERSION AL423816.1 GI:12207010
KEYWORDS GSS.
SOURCE Kluyveromyces marxianus
ORGANISM Kluyveromyces marxianus
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;

```

```

Saccharomycetales; Saccharomycetaceae; Kluyveromyces.
1 (bases 1 to 595)
Souciet,J.L., Aigle,M., Artiguenave,F., Blandin,G.,
Boloitin-Fukuhara,M., Bon,E., Brottier,P., Casaregola,S.,
de-Montigny,J., Dujon,B., Durrens,P., Lepingle,A., Llorente,B.,
Malpertuy,A., Neuvéglise,C., Ozier-Kalogeropoulos,O., Potier,S.,
Saurin,W., Tekala,F., Toffano-Nioche,C., Wesolowski-Louvel,M.,
Wincker,P. and Weissenbach,J.
Genomic exploration of the hemiascomycetous yeasts: 1. A set of
yeast species for molecular evolution studies
FEBS Lett. 487 (1), 3-12 (2000)
20584711
MEDLINE
11152876
PUBMED
2 (bases 1 to 595)
Llorente,B., Malpertuy,A., Blandin,G., Artiguenave,F., Wincker,P.
and Dujon,B.
Genomic exploration of the hemiascomycetous yeasts: 12.
Kluyveromyces marxianus var. marxianus
FEBS Lett. 487 (1), 71-75 (2000)
20584722
PUBMED
11152887
PUBMED
3 (bases 1 to 595)
Genoscope.
Direct Submission
Submitted (08-SEP-2000) Genoscope - Centre National de Sequencage,
2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :
seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
This GSS is part of a random genomic sequencing program of thirteen
yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces
exiguus, Saccharomyces servazii, Zygosaccharomyces rouxii,
Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces
lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia
angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila,
Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to
5 kb were prepared and both extremities were sequenced. See
keywords for description of this sequence and for the sequence of
the other extremity of this insert.
FEATURES
source
1. 595
/organism="Kluyveromyces marxianus"
/mol_type="genomic DNA"
/strain="CBS 712"
/variety="marxianus"
/db_xref="taxon:4911"
/clone="AZ0AA010F07"
/clone_lib="AZ0AA"
/notes="end : T7"
misc_feature
<24..>581
/notes="similar to CAB08672 [ murD ]
2 putative frameshift(s)"
/evidence=not_experimental
BASE COUNT 102 a 202 c 175 g 106 t 10 others
ORIGIN
Alignment Scores:
Pred. No.: 9.46e-11 Length: 595
Score: 187.00 Matches: 63
Percent Similarity: 48.45% Conservative: 31
Best Local Similarity: 32.47% Mismatches: 84
Query Match: 8.27% Indels: 18
DB: 29 Gaps: 3
US-09-701-229-2 (1-448) x CNS070MQ (1-595)
Qy 264 ProValGlyGluLeuLysIleArgGlyAlaHisAsnTyrSerAsnAlaLeuAlaLeu 283
Db 24 CCACCTCGATGACGTCGCTCTTCGCCCCACACATGTCGAGAACGCTTGGGCGAGT 83
Qy 284 AlaLeuGlyHisAlaValGlyLeuPropheAspAlaMetLeuGlyAlaLeuLysAlaPhe 303
Db 84 GCCTGTACTCGCTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTGCGCGGACCTC 143
Qy 304 SerGlyLeuAlaHisArgCysGlnTrpValArgGluArgGlnGlyValSerTyrTyrAsp 323

```

```

Db      144  CACCTTGGCGGCGATGTCGAGACCGTCCATAAGCTGGTGGCATCACCTGGGTCGAC 203
QY      324  AspSerLysAlaThrAsnValGlyAlaAlaLeuAlaAlaIleGluGlyLeuGlyAlaAsp 343
Db      204  GACTCCAGCGCCACCAACCTCAGCGCGGAACCTCTCGATCGGCGCTTC----- 254
QY      344  IleAspGlyLysLeuValLeuLeuAlaGlyGlyAspGlyLysGlyAlaAspPheHisAsp 363
Db      255  -----GAGCACATCGTGTGGATMGCGCGAGCAG -GCCAAGGGCCACCACTTTGACGAC 307
QY      364  Leu-ArgGluProValAlaAlaArgPheCys-ArgAlaValValLeuLeuGlyArgAspAlaG 383
Db      308  CTGGTCACACCGCCGCGAGAGCTGCGGAGCATCGTG---CTGGGAACCGACCGTG 364
QY      383  LylLeuIleAlaGlnAlaLeuGlyAsnAlaValProLeuValArgValAlaThrLeuAspG 403
Db      365  ACGTCKTCGCCAATTCITTCCTCGAGCAGCGCCGCAATTCCTCCGTCGTCATCGGATG 424
QY      403  lu-----AlaValArgGlnAlaAlaGluLeuAlaArgGluG 415
Db      425  ACATTTNGGAGAGATGTCATGGCGAGGAGTNCGGAGGCTCCGAGATGGCCCGCG 484
QY      415  LysAlaValLeuLeuSerProAlaCysAlaSerLeuAspMetPheLysAsnPheGluG 435
Db      485  GCGACACCGCTCTGTATGCTCCGGGTGCGCGACATC-GACATTTGGCGTGGATATGCTG 543
QY      435  luArgGlyArgLeuPheAlaLysAlaValGluGluLeu 447
Db      544  CCCGGGTGACGATTCGTCAACGGCGGCACGACAAATT 581

RESULT 13
BH383040/c
LOCUS
DEFINITION BH383040.1 616 bp DNA linear GSS 10-DEC-2001
            AG-ND-127115.TF ND-TAM Anopheles gambiae genomic clone AG-ND-127115
            , genomic survey sequence.
ACCESSION BH383040.1 GI:17329182
VERSION
KEYWORDS
SOURCE
ORGANISM
            Anopheles gambiae (African malaria mosquito)
            Anopheles gambiae
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea;
            Anopheles.
REFERENCE 1 (bases 1 to 616)
AUTHORS Hong, Y.S., Hogan, J.R., Wang, X., Sarkar, A., Sim, C., Loftus, B.J., Ren
            , C., Huff, E.R., Carlile, J.L., Black, K., Zhang, H.-B., Gardner, M.J.
            and Collins, F.H.
            Construction of a BAC library and generation of BAC end
            sequence-tagged connectors for genome sequencing of the African
            malaria mosquito Anopheles gambiae
            Mol. Genet. Genomics 268 (6), 720-728 (2003).
            22542063
            12655398
            Other GSSs: AG-ND-127115.TF
            Contact: Brendan J Loftus
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0208
            Fax: 301 838 3543
            Email: bjloftus@tigr.org
            This clone is from an A. gambiae BAC library (ND-TAM) provided by
            F.H. Collins and sequenced by The Institute for Genomic Research
            (TIGR). The BAC library was generated from A. gambiae PEST strain
            DNA. All DNA was extracted from newly hatched first instar larvae
            to minimize the inclusion of DNA from microorganisms that inhabit
            the gut. The DNA is derived from mixed sexes of larvae. The BAC
            library was constructed at Texas A&M University BAC Center
            University, College Station, Texas 77843-2123, USA using a HindIII
            partial digest.
            Seq primer: M13 Rev
            Class: BAC ends.

```

```

FEATURES
            source
            Location/Qualifiers
            1..616
            /organism="Anopheles gambiae"
            /mol_type="genomic DNA"
            /strain="PEST"
            /db_xref="taxon:7165"
            /clone_lib="AG-ND-127115"
            /note="Vector: pECBAC1; Site_1: HindIII"
BASE COUNT 182 a 134 c 87 g 213 t
ORIGIN
Alignment Scores:
Pred. No.: 1.3e-10 Length: 616
Score: 186.00 Matches: 46
Percent Similarity: 50.34% Conservative: 28
Best Local Similarity: 31.23% Mismatches: 69
Query Match: 8.23% Indels: 4
DB: 28 Gaps: 1
US-09-701-229-2 (1-448) x BH383040 (1-616)
QY 250 AspGlyGlnLysTrpLeuAlaPheGlnPheAspLysLeuLeuProValGlyGluLeuLys 269
Db 471 GACAATGAATTAGTTCATTAAACTTCAGGACGACTTCAGAATGAAAAATTTCTGATCTTCA 412
QY 270 IleArgGlyAlaHisAsnTy-SerAsnAlaLeuAlaAlaLeuAlaLeuGlyHisAlaVal 289
Db 411 CTGTGAGGAACCATTAATGTTCCCAATAGTTAGCTGCAATATAGCAGGTAAATCGCTA 352
QY 290 GlyLeuProPheAspAlaMetLeuGlyAlaLeuLysAlaPheSerGlyLeuAlaHisArg 309
Db 351 AATATTAGTAATGAAGATATCGCAATAGTCTGATGACATTTTCAGGCTGTTCCGCACAGA 292
QY 310 CysGlnTrpValArgGluArgGlnGlyValSerTy-TyrAspAspSerLysAlaThrAsn 329
Db 291 TTGGAGCAGGTTGCAGTATTAAATGATGTAATACATCAATGCAATGCAAGCAACCAAT 232
QY 330 ValGlyAlaAlaLeuAlaAlaIleGluGlyLeuGlyAlaAspIleAspGlyLysLeuVal 349
Db 231 GTAAATGCTGCATATTATGCACTGGAA-----AGTGAATAATATCTGCAATA 184
QY 350 LeuLeuAlaGlyGlyAspGlyLysGlyAlaAspPheHisAspLeuArgGluProValAla 369
Db 183 TGGATTGTTGGTGGAAATAGATAAGGAATGACTATACAGAAATTCAGGATCTCGTAAG 124
QY 370 ArgPheCysArgAlaValValLeuLeuGlyArgAspAlaGlyLeuIleAlaGlnAlaLeu 389
Db 123 AAGAAAGTTCGTGCCATCGTATGCTGGGATGGGATTAATGAAAGATCATTCAGTCTTC 64
QY 390 GlyAsnAlaValProLeuVal 396
Db 63 AGAAACAAAAAGGATTGATT 43
RESULT 14
BH369957/c
LOCUS
DEFINITION BH369957 495 bp DNA linear GSS 10-DEC-2001
            AG-ND-138K1.TF ND-TAM Anopheles gambiae genomic clone AG-ND-138K1,
            genomic survey sequence.
ACCESSION BH369957
VERSION
KEYWORDS
SOURCE
ORGANISM
            Anopheles gambiae (African malaria mosquito)
            Anopheles gambiae
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea;
            Anopheles.
REFERENCE 1 (bases 1 to 495)
AUTHORS Hong, Y.S., Hogan, J.R., Wang, X., Sarkar, A., Sim, C., Loftus, B.J., Ren
            , C., Huff, E.R., Carlile, J.L., Black, K., Zhang, H.-B., Gardner, M.J.
            and Collins, F.H.
            Construction of a BAC library and generation of BAC end
            sequence-tagged connectors for genome sequencing of the African

```

```

JOURNAL      malaria mosquito Anopheles gambiae
MEDLINE      Mol. Genet. Genomics 268 (6), 720-728 (2003)
PUBMED      22542063
COMMENT      12655398
              Contact: Brendan J Loftus
              Department of Eukaryotic Genomics
              The Institute for Genomic Research
              9712 Medical Center Dr., Rockville, MD 20850, USA
              Tel: 301 838 0208
              Fax: 301 838 3543
              Email: bjloftus@tigr.org
              This clone is from an A. gambiae BAC library (ND-TAM) provided by
              F.H. Collins and sequenced by The Institute for Genomic Research
              (TIGR). The BAC library was generated from A. gambiae PEST strain
              DNA. All DNA was extracted from newly hatched first instar larvae
              to minimize the inclusion of DNA from microorganisms that inhabit
              the gut. The DNA is derived from mixed sexes of larvae. The BAC
              library was constructed at Texas A&M University BAC Center
              University, College Station, Texas 77843-2123, USA using a HindIII
              partial digest.
              Seq primer: M13 For
              Class: BAC ends.

FEATURES             source
    Location/Qualifiers
        1..495
            /organism="Anopheles gambiae"
            /mol_type="genomic DNA"
            /strain="PEST"
            /db_xref="taxon:7165"
            /clone="AG-ND-138K1"
            /clone_lib="ND-TAM"
            /note="vector: pECBAC1; Site_1: HindIII"
BASE COUNT      143 a 110 c 68 g 174 t
ORIGIN
Alignment Scores:
Pred. No.:      3.15e-09      Length:      495
Score:          172.50      Matches:      43
Percent Similarity: 52.24%      Conservative: 27
Best Local Similarity: 32.09%      Mismatches: 59
Query Match:    7.63%      Indels:      5
DB:             28      Gaps:      2

US-09-701-229-2 (1-448) x BH369957 (1-495)
Qy 274 HisAsnTyrSerAsnAlaLeuAlaLeuAlaLeuGlyHisAlaValGlyLeuProPhe 293
||||| :||||:||||| :||| :||| :|||
Db 392 CATAATGTGCAATAGTTTGTGCTGCAAGTATACGAGTAACTGCTAAATATTAGTAAT 333
||||| :||||:||||| :||| :||| :|||
Qy 294 AspAlaMetLeuGlyAlaLeuLysAlaPheSerGlyLeuAlaHisArgCysGlnTrpVal 313
||||| :||||:||||| :||| :||| :|||
Db 332 GAAAGTATCGCAATAGCTGATGACATTTTCAGGCTGTCGCGACAGATTGGACAGGTT 273
||||| :||||:||||| :||| :||| :|||
Qy 314 ArgGluArgGlnGlyValSerTyrTyrAspSerLysAlaThrAsnValGlyAlaAla 333
||||| :||||:||||| :||| :||||:||||| :||||| :|||||
Db 272 GCAGTTATTATGATGTAATGAAATCATCATGACAGTAAAGCAACCAATGTAATGCTGCA 213
||||| :||||:||||| :||| :||||:||||| :||||| :|||||
Qy 334 LeuAlaAlaLeuGlyLeuGlyAlaAspIleAspGlyLysLeuValLeuLeuAlaGly 353
||||| :||||:||||| :||| :||| :|||
Db 212 TATTATGACGTGAA-----AGTGTGAATATCTCGCAATATGATTTGTTGGT 165
||||| :||||:||||| :||| :||| :|||
Qy 354 GlyAspGlyLysGlyAlaAspPheHisAspLeuArgGluProValAlaAa-gPheCysArg 373
||||| :||||:||||| :||| :||| :|||
Db 164 GGAATAGATAAGGAAATGACTATACAGAAATTCAGGATCTGGTAAAGAAAGTTCGT 105
||||| :||||:||||| :||| :||| :|||
Qy 374 AlaValValLeuLeuGlyArgAspAlaGlyLeuIleAlaGlnAlaLeuGlyAsnAlaVal 393
||||| :||||:||||| :||| :||| :|||
Db 104 GCCATCGTATGCTCGGATGGAATAATGAAAGATCATTCAGTTCTTCAGAAACAAAAG 45
||||| :||||:||||| :||| :||| :|||
Qy 394 ProLeuVal---ArgValAlaThrLeuAspGluAlaValArg 406
||||| :||||:||||| :||| :||| :|||
Db 44 GATTGTATTTTGAAACCTCTAGCATGGAAGATGTTGTAAG 3
||||| :||||:||||| :||| :||| :|||
RESULT 15

```

```

BH382726/c
LOCUS      BH382726      378 bp      DNA      linear      GSS 10-DEC-2001
DEFINITION AG-ND-127B23_TR ND-TAM Anopheles gambiae genomic clone AG-ND-127B23
            , genomic survey sequence.
ACCESSION  BH382726
VERSION    BH382726.1 GI:17328868
KEYWORDS   GSS.
SOURCE     Anopheles gambiae (African malaria mosquito)
ORGANISM   Anopheles gambiae
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea;
            Anopheles.
REFERENCE  1 (bases 1 to 378)
AUTHORS   Hong,X.S., Hogan,J.R., Wang,X., Sarkar,A., Sim,C., Loftus,B.J., Ren
            ,C., Huff,E.R., Carlile,J.L., Black,K., Zhang,H.-B., Gardner,M.J.
            and Collins,F.H.
TITLE     Construction of a BAC library and generation of BAC end
            sequence-tagged connectors for genome sequencing of the African
            malaria mosquito Anopheles gambiae
JOURNAL    Mol. Genet. Genomics 268 (6), 720-728 (2003)
MEDLINE    22542063
PUBMED     12655398
COMMENT    Contact: Brendan J Loftus
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0208
            Fax: 301 838 3543
            Email: bjloftus@tigr.org
            This clone is from an A. gambiae BAC library (ND-TAM) provided by
            F.H. Collins and sequenced by The Institute for Genomic Research
            (TIGR). The BAC library was generated from A. gambiae PEST strain
            DNA. All DNA was extracted from newly hatched first instar larvae
            to minimize the inclusion of DNA from microorganisms that inhabit
            the gut. The DNA is derived from mixed sexes of larvae. The BAC
            library was constructed at Texas A&M University BAC Center
            University, College Station, Texas 77843-2123, USA using a HindIII
            partial digest.
            Seq primer: M13 Rev
            Class: BAC ends.

FEATURES             source
    Location/Qualifiers
        1..378
            /organism="Anopheles gambiae"
            /mol_type="genomic DNA"
            /strain="pEST"
            /db_xref="taxon:7165"
            /clone="AG-ND-127B23"
            /clone_lib="ND-TAM"
            /note="vector: pECBAC1; Site_1: HindIII"
BASE COUNT      108 a 85 c 52 g 133 t
ORIGIN
Alignment Scores:
Pred. No.:      2.38e-09      Length:      378
Score:          172.00      Matches:      42
Percent Similarity: 52.00%      Conservative: 23
Best Local Similarity: 33.60%      Mismatches: 56
Query Match:    7.61%      Indels:      4
DB:             28      Gaps:      1

US-09-701-229-2 (1-448) x BH382726 (1-378)
Qy 267 GluLeuLysIleArgGlyAlaHisAsnTyrSerAsnAlaLeuAlaLeuAlaLeuGly 286
||||| :||||:||||| :||| :||||:||||| :||||| :|||||
Db 375 GATCTTCACATTTAGGAAACCAATATGTTGCCAATAGTTTACGTCGAAGTATACGAGT 316
||||| :||||:||||| :||| :||||:||||| :||||| :|||||
Qy 287 HisAlaValGlyLeuProPheAspAlaMetLeuGlyAlaLeuLysAlaPheSerGlyLeu 306
||||| :||||:||||| :||| :||| :|||
Db 315 AAACGTCTAAATATTAGTAATGAAAGTATCCGCAATGCTGATGACATTTTCAGGCTGTT 256
||||| :||||:||||| :||| :||| :|||
Qy 307 AlaHisArgCysGlnTrpValArgGluArgGlnGlyValSerTyrTyrAspSerLys 326
||||| :||||:||||| :||| :||| :|||
Db 255 CCGCACAGATTGGACAGGTTGCAGTTATTATGATGTGGAATACATCATCATGACAGTAAG 196
||||| :||||:||||| :||| :||| :|||

```

Search completed: August 14, 2003, 09:10:32  
Job time : 2302 secs

; TYPE: nucleic acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: mRNA  
 ; HYPOTHETICAL: NO  
 ; ANTI-SENSE: NO  
 ; US-08-843-309-3

## Alignment Scores:

Pred. No.: 1.86e-55 Length: 1350  
 Score: 586.50 Matches: 162  
 Percent Similarity: 52.92% Conservativity: 83  
 Best Local Similarity: 34.99% Mismatches: 177  
 Query Match: 25.95% Indels: 41  
 DB: 2 Gaps: 14

US-09-701-229-2 (1-448) x US-08-843-309-3 (1-1350)

```

Qy 7 AspHisPheArg-----IleValValGlyLeuGlyLysSerGlyMetSerLeu 22
Db 13 GAUCAUUUAAAAAAGAAAGUUCUUUAGUUUGGCCAAGUCUGGUAACUGCA 72

Qy 23 ValArgTyrLeuAlaArgArgGlyLeuProPheAlaValValAspThrArg----- 39
Db 73 GCUCGUUUUGUUGCAAGCAGGUGGCCAUGUGACAGUAAAGUAGGAAACCUUUCGAG 132

Qy 40 GluAsnProGluLeuLeuAlaThrLeuArgAlaGlnTyrProGlnValGluValArgCys 59
Db 133 GACAAUCCAGCUGCCCAAGUUUCUGGAA-----GAGGGAUCCAGGUAUACA 183

Qy 60 GlyGlu-----LeuAspAlaGluPheLeuCysSerAlaArgGluLeuTyr 74
Db 184 GGUGCCCAUCCUUUGGAACUCUUGGAAGAUUUGCCUUUAGUGGAAAA----- 234

Qy 75 ValSerProGlyLeuSerLeuArgThrProAlaLeuValGlnAlaAlaLysGlyVal 94
Db 235 ---AAUCCAGUAUCCCUCAACAAUCCAUCCAUCCAUCCAUCCAUCCAUCCAUCCAUCC 291

Qy 95 ArgIleSerGlyAspIleAspLeuPheAlaArgGluAlaLysAlaProIleValAlaIle 114
Db 292 CCAGUCUUGACUGAGGUGGAUUGGUUUUUGAUUUGAUUUGAUUUGAUUUGAUUUGAUUUG 351

Qy 115 ThrGlySerAsnAlaLysSerThrValThrLeuValGlyGluMetAlaValAlaAla 134
Db 352 ACAGGAUCCAGGUAAGAACACCAACACGACAUUAGUUGGGAAGUUUUGACUGCGCU 411

Qy 135 AspLysArgValAlaValGlyGlyAsnLeuGlyThrProAlaLeuAspLeu-----Leu 152
Db 412 GGGCAACAUGGUCUUUUUAUCCAGGGAUUAUCCGCUUUAUCCGCGCAGUUGGCUCAAUA 471

Qy 153 AlaAspAspIleGluLeuTyrValLeuGluLeuSerSerPheGlnLeuGluThrCysasp 172
Db 472 GCAUCAGAAAGGACAGCGUUGUUGAACUUUUCUUUUCUUUUCUUUUCUUUUCUUUUCUU 531

Qy 173 ArgLeuAsnAlaGluValAlaThrValLeuAsnValSerGluAspHisMetAspArgTyr 192
Db 532 GAUUUCCAUCCAGAGAUUGCGGUUUUAUCCAAUCCAUCCAUCCAUCCAUCCAUCCAUCCAU 591

Qy 193 AspGlyMetAlaAspTyrHisLeuAlaLysHisArgIle-----PheArgGlyAla 209
Db 592 GGGUUAUUUUGGAUUAUUGAGCAGCCCAAGGUAUUAUCCAGAAUUAUCCAGAAUUAUCCAG 651

Qy 210 ArgGlnValValAlaAsnArgAlaAspAlaLeuThrArgProLeuIleAlaAspThr--- 228
Db 652 GAUUUCCUUGUUGUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUA 711

Qy 229 -----ValProCysTyrSerPheGlyLeuAsnLysProAspPheLysAlaPheGly 245
Db 712 GCCACUGUUGUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUA 753

Qy 246 LeuIleGluGluAspGlyGlnLysTrpLeuAlaPheGlnPheAspLysLeuLeuProVal 265
Db 754 GCUUAUCCUGGAAGAUGGUCAA-----CUCUACUCCUGGUGGAAGUAGUAGUAGGCGCG 807
  
```

```

Qy 266 GlyGluLeuLysIleArgGlyAlaHisAsnTyrSerAsnAlaLeuAlaLeuAlaLeu 285
Db 808 AAUGAAACUGGUGUCCAGUAGCCACAAGUGGAAAGAAUGCCUUGGCGAUUUGUGUA 867

Qy 286 GlyHisAlaValGlyLeuProPheAspAlaMetLeuGlyAlaLeuLysAlaPheSerGly 305
Db 868 GCCAAGCUGUGAUGGACAAUCAAACCAACCAAGGAAACUUAUCCAGGCUUGGUGGU 927

Qy 306 LeuAlaHisArgCysGlnTrpValArgGluArgGlnGlyValSerTyrTyrAspSer 325
Db 928 GUCAACACCGCUGUCCAGUUUGUGGAGUAGCAUCAAAGGUGUUAAAUUUAUUAUUAU 987

Qy 326 LysAlaThrAsnValGlyAlaAlaLeuAlaIleGluGlyLeuGlyAlaAspIleasp 345
Db 988 AAUCAACUAUAUCCUGGCUACUCAAACCAACCAAGGAAACUUAUCCAGGAAUUGAC 1038

Qy 346 GlyLysLeuValLeuLeuAlaGlyGlyAspGlyLysGlyAlaAspPheHisAspLeuArg 365
Db 1039 AGCAAGGCGUCUUGAUGGAGGUGUUGGAGGUGGCGGCAAGUAGUUAUUGAGCAAUUG 1095

Qy 366 GluProValAlaArgPheCysArgAlaValValLeuLeuGlyArgAspAlaGlyLeuIle 385
Db 1096 GUGCCAGACAUUCCUGGACUCAAAGAGUAGUCCUUGGUGUUAUCCAGGAGGUGUUC 1155

Qy 386 AlaGlnAlaLeuGlyAsnAla---ValProLeuValArgValAlaThrLeuAspGluAla 404
Db 1156 AAACGGGCGACAGACAGGCGUGGUGUCCUUAUUGGAGGCGCAGAGAUUUGCAGAUCCG 1215

Qy 405 ValArgGlnAlaAlaGluLeuAlaArgGluGlyAspAlaValLeuLeuSerProAlaCys 424
Db 1216 ACCGCAAGGCGUAGAGUAGGUGGACUACAGGAGAGUAGUUGUUCUUAUUGUCCGCAU 1275

Qy 425 AlaSerLeuAspMetPheLysAsnPheGluGluArgGlyArgLeuPheAlaLysAlaVal 444
Db 1276 GCCAGCUGGGAUAGUAGUAGUAGUAGUAGUAGUAGUAGUAGUAGUAGUAGUAGUAGU 1335

Qy 445 GluGluLeu 447
Db 1336 GCGGAGUUA 1344
  
```

## RESULT 15

```

US-08-843-309-1
; Sequence 1, Application US/08843309
; Patent No. 5834270
; GENERAL INFORMATION:
; APPLICANT: Skatrud, Paul
; APPLICANT: Peery, Robert
; APPLICANT: Hoskins, JoAnn
; APPLICANT: Wu, Chyun-Yeh Earnest
; TITLE OF INVENTION: Biosynthetic Gene Mur D of Streptococcus pneumoniae
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Eli Lilly and Company
; STREET: Lilly Corporate Center
; CITY: Indianapolis
; STATE: Indiana
; COUNTRY: US
; ZIP: 46285
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION NUMBER: US/08/843,309
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Webster, Thomas D
; REGISTRATION NUMBER: 39,872
; REFERENCE/DOCKET NUMBER: X-9900
; TELECOMMUNICATION INFORMATION:
  
```

TELEPHONE: 317-276-3334  
 TELEFAX: 317-276-3861  
 INFORMATION FOR SEQ ID NO: 1:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 1353 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: DNA (genomic)  
 HYPOTHETICAL: NO  
 ANTI-SENSE: NO  
 FEATURE:   
 NAME/KEY: CDS  
 LOCATION: 1..1353  
 US-08-843-309-1

*Sequence!*

*US Pat. 5,834,270*

Alignment Scores:  
 Pred. No.: 1 87e-55 Length: 1353  
 Score: 586.50 Matches: 162  
 Percent Similarity: 52.92% Conservative: 83  
 Best Local Similarity: 34.99% Mismatches: 177  
 Query Match: 25.95% Indels: 41  
 DB: 2 Gaps: 14

US-09-701-229-2 (1-448) x US-08-843-309-1 (1-1353)

QY 7 AspHisPheArg-----lleValValGlyLeuGlyLysSerGlyMetSerLeu 22  
 DB 13 GATCAATTAATAAAGAAAGTTCTTTAGTTTGGCCCAAGTCTGGTGAATCTGCA 72  
 QY 23 ValArgTyrLeuAlaArgGlyLeuProPheAlaValValAspThrArg----- 39  
 DB 73 GCTCGTTTGTGGACAAAGTAGTGCCATTGTGACAGTAATGATGGGAACCTTTCGAG 132  
 QY 40 GluAsnProGluLeuAlaThrLeuArgAlaGlnTyrProGlnValGluValArgCys 59  
 DB 133 GACAATCCAGCTCCCAAGATTGCTGGAA-----GAAGGGATCAAGTCAATACA 183  
 QY 60 GlyGlu-----LeuAspAlaGluPheLeuCysSerAlaArgGluLeuTyr 74  
 DB 184 GGTGGCCATCCTTTGGAACCTTTGGATGAAGAGTTTGCCTTATGCTGAAA----- 234  
 QY 75 ValSerProGlyLeuSerLeuArgThrProAlaLeuValGlnAlaAlaLysGlyVal 94  
 DB 235 ---ATTCAGGTATCCCTACAAATCCATGATGAAAGGCTTGGCCCAAGAGAATT 291  
 QY 95 ArgIleSerGlyAspIleAspLeuPheAlaArgGluAlaLysAlaProIleValAlaIle 114  
 DB 292 CCAGTCTTGAAGTGAAGTGGCTTATTTGATTTTCAAGACCCGATTTATGATATC 351  
 QY 115 ThrGlySerAsnAlaLysSerThrValThrLeuValGlyGluMetAlaValAla 134  
 DB 352 ACAGGATCGAAGCGGTAAGAACCAACCAAGCTATGATTTGGGGAAGTTTGAAGTCTGCT 411  
 QY 135 AspLysArgValAlaValGlyGlyAsnLeuGlyThrProAlaLeuAspLeu-----Leu 152  
 DB 412 GGGCAACATGGTCTTTATCAGGGAATATCGGCTATCTGCGCAGTCAGTCTCTCAATA 471  
 QY 153 AlaAspAspIleGluLeuTyrValLeuGluLeuSerSerPheGlnLeuGluThrCysAsp 172  
 DB 472 GCATCAGATAAGGACACGCTTCTTATGGAACCTTCTTCTTCAACTCATGGTGTTCAA 531  
 QY 173 ArgLeuAsnAlaGluValAlaThrValLeuAsnValSerGluAspHisMetAspArgTyr 192  
 DB 532 GAATTCATCCAGATGTCGGTTATTACCAACCTCATGCCAACCTCATATCGACTACCAT 591  
 QY 193 AspGlyMetAlaAspTyrHisLeuAlaLysHisArgIle-----PheArgGlyAla 209  
 DB 592 GGGTCATTTCTGAATATGTAGCAGCAAGTGGGAATATCCAGAACCAAGATGACACAGCT 651  
 QY 210 ArgGlnValValAlaAsnArgAlaAspAlaLeuThrArgProLeuIleAlaAspThr--- 228  
 DB 652 GATTCCTTGTCTTGAACCTTTAATCAAGACTTGGCAAAAGACTTCACTTCCCAAGACAA 711

QY 229 -----ValProCysTrpSerPheGlyLeuAsnLysProAspPheLysAlaPheGly 245  
 DB 712 GCCACTGTTGTACCACTTTTCAACA-----CTTGAAGAGTTGAT-----GGA 753  
 QY 246 LeuIleGluGluAspGlyGlnLysTrpLeuAlaPheGlnPheAspLysLeuLeuProVal 265  
 DB 754 GCTTATCTGGAAGATGGTCAA-----CTCTACTTCGTTGGTGAAGTAGTATGCGCAGCG 807  
 QY 266 GlyGluLeuLysIleArgGlyAlaHisAsnTyrSerAsnAlaLeuAlaAlaLeuAlaLeu 285  
 DB 808 AATGAATCGGTGTTCCAGGATAGCCACAATGTGGAATATGCCCTTCGACTATTGCTGTA 867  
 QY 286 GlyHisAlaValGlyLeuProPheAspAlaMetLeuGlyAlaLeuLysAlaPheSerGly 305  
 DB 868 GCCAAGCTTCGTGATGTGGACAATCAACCAATCAAGGAACCTTTTCACCCCTCGGTGT 927  
 QY 306 LeuAlaHisArgCysGlnTrpValArgGluArgGlnGlyValSerTyrTyrAspAspSer 325  
 DB 928 GTCAACACCCGCTCTCCAGTTTGTGGATGATCAACAGGGTGTAAATTTCTATAACGACAGT 987  
 QY 326 LysAlaThrAsnValGlyAlaAlaLeuAlaIleGluGlyLeuGlyAlaAspIleAsp 345  
 DB 988 AATCAACTAATATCTTGGCTACTCAAAAGGCTTATCAGGATTGAC-----AAC 1038  
 QY 346 GlyLysLeuValLeuLeuAlaGlyAspGlyLysGlyAlaAspPheHisAspLeuArg 365  
 DB 1039 AGCAAGGTGCTCTTGTATGTCAGGTGTTGGACCGTGGCAATGAGTTTCAAGAAATTG--- 1095  
 QY 366 GluProValAlaArgPheCysArgAlaValValLeuLeuGlyArgAspAlaGlyLeuIle 385  
 DB 1096 GTGCCAGACATTAATGAGTCAAGAGATGGTCACTCTGGTCAATCTCGCAACGCTGTC 1155  
 QY 386 AlaGlnAlaLeuGlyAsnAla---ValProLeuValArgValAlaThrLeuAspGluAla 404  
 DB 1156 AAACGGCAGCAGACAGAGCTGCTGCTTATGTTGGAGCCGACAGATATTCAGATGCG 1215  
 QY 405 ValArgGlnAlaAlaGluLeuAlaArgGluGlyAspAlaValLeuLeuSerProAlaCys 424  
 DB 1216 ACCCGCAAGGCCCTATGAGCTTCGACTCAAGGAGATGTGTTCTTCTTCTTCTGCTCAAT 1275  
 QY 425 AlaSerLeuAspMetPheLysAsnPheGlyGluGlyArgGlyArgLeuPheAlaLysAlaVal 444  
 DB 1276 GCCACTGGATATGATGCTAATCTTGAAGTACGTGGCGACCTCTTTATCGACACAGTA 1335  
 QY 445 GluGluLeu 447  
 DB 1336 GCGGAGTTA 1344

Search completed: August 14, 2003, 09:37:52  
 Job time : 1636 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run On: August 14, 2003, 04:12:30 ; Search time 2898 seconds

(without alignments)  
12160.625 Million cell updates/sec

Title: US-09-701-229-1

Perfect score: 1450

Sequence: 1 cgtgctgacggcctcgcca.....tggtgagcggcgacggcgtac 1450

Scoring table: OLIGO\_NUC

Gapop 60.0 , Gapext 60.0

Searched: 22781392 seqs, 12152238056 residues

Word size : 0

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

EST:\*

- 1: em\_estba:\*\*
- 2: em\_esthum:\*\*
- 3: em\_estin:\*\*
- 4: em\_estmu:\*\*
- 5: em\_estov:\*\*
- 6: em\_estpl:\*\*
- 7: em\_estro:\*\*
- 8: em\_htc:\*\*
- 9: gb\_est1:\*\*
- 10: gb\_est2:\*\*
- 11: gb\_htc:\*\*
- 12: gb\_est3:\*\*
- 13: gb\_est4:\*\*
- 14: gb\_est5:\*\*
- 15: em\_estfun:\*\*
- 16: em\_estom:\*\*
- 17: em\_gss\_hum:\*\*
- 18: em\_gss\_inv:\*\*
- 19: em\_gss\_pin:\*\*
- 20: em\_gss\_vrt:\*\*
- 21: em\_gss\_fun:\*\*
- 22: em\_gss\_mam:\*\*
- 23: em\_gss\_mus:\*\*
- 24: em\_gss\_pro:\*\*
- 25: em\_gss\_rod:\*\*
- 26: em\_gss\_phg:\*\*
- 27: em\_gss\_vrl:\*\*
- 28: gb\_gss1:\*\*
- 29: gb\_gss2:\*\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
c 1	600	41.4	1445	29	BZ568959
c 2	510	35.2	892	29	BZ579116
c 3	125	8.6	1171	29	BZ559806
c 4	65	4.5	1194	29	BZ550283

c 5	23	1.6	1754	10	BE966412
c 6	22	1.5	537	9	AL825430
c 7	22	1.5	687	10	BE586183
c 8	21	1.4	265	14	CB061196
c 9	21	1.4	369	9	AA969049
c 10	21	1.4	501	12	BI306480
c 11	21	1.4	792	10	BE041029
c 12	20	1.4	286	9	AW771204
c 13	20	1.4	289	14	CB042199
c 14	20	1.4	301	10	BF657011
c 15	20	1.4	318	9	AW139275
c 16	20	1.4	323	9	AW136188
c 17	20	1.4	338	13	BQ552547
c 18	20	1.4	376	10	BF109460
c 19	20	1.4	381	9	AI824553
c 20	20	1.4	397	9	AA013254
c 21	20	1.4	409	28	BH799135
c 22	20	1.4	440	10	BE633796
c 23	20	1.4	441	13	BY080930
c 24	20	1.4	472	9	AI291792
c 25	20	1.4	481	28	BH376496
c 26	20	1.4	489	13	BU040543
c 27	20	1.4	511	13	BQ286376
c 28	20	1.4	561	12	BI792784
c 29	20	1.4	571	12	BI792863
c 30	20	1.4	584	13	BU043073
c 31	20	1.4	589	12	BI792852
c 32	20	1.4	596	13	BQ269723
c 33	20	1.4	621	13	BU040487
c 34	20	1.4	633	9	AI819231
c 35	20	1.4	634	29	BX126826
c 36	20	1.4	665	9	AW191059
c 37	20	1.4	681	29	BZ781755
c 38	20	1.4	738	29	BX231314
c 39	20	1.4	768	29	CC008801
c 40	20	1.4	812	9	AL695087
c 41	20	1.4	815	29	CC008799
c 42	20	1.4	816	28	AY079596
c 43	20	1.4	911	13	BU090243
c 44	20	1.4	928	29	BZ567557
c 45	20	1.4	1085	29	CNS077DK

ALIGNMENTS

RESULT 1

BZ568959/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

BZ568959 1445 bp DNA linear GSS 17-DEC-2002  
pacs2-164\_8171.y2 pacs2-164 Pseudomonas aeruginosa genomic clone  
pacs2-164\_8171, genomic survey sequence.

BZ568959

BZ568959

BZ568959.1 GI:27202799

GSS.

Pseudomonas aeruginosa

Pseudomonas aeruginosa

Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;

Pseudomonadaceae; Pseudomonas.

1 (bases 1 to 1445)

Spencer, D.H., Raymond, C.K., Smith, E.E., Sims, E.E., Hastings, M.,

Burns, J.L., Kaul, R. and Olsen, M.V.

Whole-Genome-Sequence variation among multiple isolates of

Pseudomonas aeruginosa library

J. Bacteriol., (2002) In press

Contact: Chris K. Raymond

Genome Center

University of Washington

Box 352145, Seattle, WA 98105-2145, USA

Tel: 2062216954

Fax: 2066857244

Email: craymond@u.washington.edu

Class: shotgun.

Location/Qualifiers

FEATURES

```

source
1. 1445
/organism="Pseudomonas aeruginosa"
/mol_type="genomic DNA"
/strain="2-164"
/db_xref="taxon:287"
/clone="pacs2-164_8171"
/clone_lib="pacs2-164"
/notes="clinical isolate 2-164 Whole genomic shotgun library."
BASE COUNT 308 a 424 c 412 g 293 t 8 others
ORIGIN

Query Match 41.4%; Score 600; DB 29; Length 1445;
Best Local Similarity 100.0%; Pred. No. 9.3e-289;
Matches 600; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 488 CCGGGCGCTGACCTGCTGCGCCGACGACATCGAGCTGTAGCTGTGAGCTGTCGAGCTT 547
Db 676 CCGGGCGCTGACCTGCTGCGCCGACGACATCGAGCTGTAGCTGTGAGCTGTCGAGCTT 617

QY 548 CCAGCTGGAAACCTGGGATCGCTCAACGCCGAGGTGGCGACCGCTGCTGAACGTCAGCGA 607
Db 616 CCAGCTGGAAACCTGGGATCGCTCAACGCCGAGGTGGCGACCGCTGCTGAACGTCAGCGA 557

QY 608 AGACCATATGATGCTAGACGCGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 667
Db 556 AGACCATATGATGCTAGACGCGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 497

QY 668 CCGCGTGGCCGCGCAGTGTGCTGATCGCGCGGATGCTGACCGCGGATGCTGACCGCGGATGCTG 727
Db 496 CCGCGTGGCCGCGCAGTGTGCTGATCGCGCGGATGCTGACCGCGGATGCTGATGCTGCTG 437

QY 728 CGATACCGTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 787
Db 436 CGATACCGTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 377

QY 788 GATCGAGGAGAGCGGCCAGAGTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 847
Db 376 GATCGAGGAGAGCGGCCAGAGTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 317

QY 848 CGAACTGAAGATCCGTGGCGCCACAACTATTCGAACGGCTGCTGCGCGCTGCTGCGCGCTGG 907
Db 316 CGAACTGAAGATCCGTGGCGCCACAACTATTCGAACGGCTGCTGCGCGCTGCTGCGCGCTGG 257

QY 908 CCATGCGGTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 967
Db 256 CCATGCGGTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 197

QY 968 GCGTCATCGTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1027
Db 196 GCGTCATCGTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 137

QY 1028 GGCACCAACGTCGGCGCGCCCTGCTGCGCGATTCGAGGGGCTGGGTGCGGATCGACCG 1087
Db 136 GGCACCAACGTCGGCGCGCCCTGCTGCGCGATTCGAGGGGCTGGGTGCGGATCGACCG 77

```

```

RESULT 2
Bz579116/c
LOCUS
DEFINITION msh2_6205.y2 msh Pseudomonas aeruginosa genomic clone msh2_6205,
genomic survey sequence.
ACCESSION Bz579116
VERSION Bz579116.1 GI:27214177
KEYWORDS GSS.
SOURCE Pseudomonas aeruginosa
ORGANISM Pseudomonas aeruginosa
Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
Pseudomonadaceae; Pseudomonas.
1 (bases 1 to 892)
REFERENCE Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M.,
AUTHORS Burns,J.L., Kaul,R. and Olsen,M.V.
TITLE Whole-Genome-Sequence variation among multiple isolates of

```

```

Pseudomonas aeruginosa library
J. Bacteriol., (2002) In press
Contact: Chris K. Raymond
Genome Center
University of Washington
Box 352145, Seattle, WA 98105-2145, USA
Tel: 2062216954
Fax: 2066857244
Email: craymond@u.washington.edu
Class: shotgun.
Location/Qualifiers
1. .892
/organism="Pseudomonas aeruginosa"
/mol_type="genomic DNA"
/strain="Msh"
/db_xref="taxon:287"
/clone="msh2_6205"
/clone_lib="msh"
/notes="Environmental isolate. Whole genomic shotgun library."
BASE COUNT 173 a 277 c 276 g 160 t 6 others
ORIGIN

```

```

Query Match 35.2%; Score 510; DB 29; Length 892;
Best Local Similarity 99.8%; Pred. No. 8.4e-244;
Matches 560; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 319 CCGCGAAAGCGGTGGCGATCTCGGTGACATCGATCTTTCGCCCGCGAGGCGAAGGCC 378
Db 644 CCGCGAAAGCGGTGGCGATCTCGGTGACATCGATCTTTCGCCCGCGAGGCGAAGGCC 585

QY 379 CGATCGTCCCATCATCCGGTTCACGCGAAGAGCACCGTGCACACCTCTGGTGGCGAAA 438
Db 584 CGATCGTCCCATCATCCGGTTCACGCGAAGAGCACCGTGCACACCTCTGGTGGCGAAA 525

QY 439 TGGCGGTGGCGGAGAACAGCGTGTGCGCGTGGCGGCAACCTCGCGACCCCGCGCGCTCG 498
Db 524 TGGCGGTGGCGGAGAACAGCGTGTGCGCGTGGCGGCAACCTCGCGACCCCGCGCGCTCG 465

QY 499 ACTGCTGTCGCGACGACATCGAGCTGTAGCTGTGGAGCTGTGAGCTTCCAGCTCGAAA 558
Db 464 ACTGCTGTCGCGACGACATCGAGCTGTAGCTGTGGAGCTGTGAGCTTCCAGCTCGAAA 405

QY 559 CTTGCGATCCCTCAACGCCGAGGTGGCGACCGTGTGAACGTCAGCGAAGACCATATGG 618
Db 404 CTTGCGATCCCTCAACGCCGAGGTGGCGACCGTGTGAACGTCAGCGAAGACCATATGG 345

QY 619 ATCGCTACGACGCGATGGCTGACTACCACTGGCCCAAGCACCGGATCTTCCGCGGTGCC 678
Db 344 ATCGCTACGACGCGATGGCTGACTACCACTGGCCCAAGCACCGGATCTTCCGCGGTGCC 285

QY 679 GCGAGTCTGTGTGAATCGCGCGATGCGCTGACCGCGACCGCTGATCGCGATACCGTGC 738
Db 284 GCGAGTCTGTGTGAATCGCGCGATGCGCTGACCGCGACCGCTGATCGCGATACCGTGC 225

QY 739 CCGTGTGTCGTCGCGCTGAAACAGCCGAGCTTCAAGGCTTTCGCGCTGTGCGAAG 798
Db 224 CCGTGTGTCGTCGCGCTGAAACAGCCGAGCTTCAAGGCTTTCGCGCTGTGCGAAG 165

QY 799 ACGGCGAAGTGGCTGGCGTTCCAGTTTCGAGTTTCGCGGTTGGCGAAGTGAAGA 858
Db 164 ACGGCGAAGTGGCTGGCGTTCCAGTTTCGAGTTTCGCGGTTGGCGAAGTGAAGA 105

QY 859 TCCGTGGCGGCCACCAACTATT 879
Db 104 TCCGTGGCGGCCACCAACTATT 84

```

```

RESULT 3
Bz559806
LOCUS
DEFINITION pacs2-164_1863.xl pacs2-164 Pseudomonas aeruginosa genomic clone
pacs2-164_1863, genomic survey sequence.
1171 bp DNA linear GSS 17-DEC-2002
Bz559806

```

ACCESSION	B2550283
VERSION	B2550283.1
KEYWORDS	GI:27153864
SOURCE	GSS.
ORGANISM	<i>Pseudomonas aeruginosa</i>
REFERENCE	<i>Pseudomonas aeruginosa</i> Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales; Pseudomonadaceae; <i>Pseudomonas</i> . 1 (bases 1 to 1194)
AUTHORS	Spencer, D.H., Raymond, C.K., Smith, E.E., Sims, E.E., Hastings, M., Burns, J.L., Kaul, R. and Olsen, M.V.
TITLE	Whole-Genome-Sequence Variation among multiple isolates of <i>Pseudomonas aeruginosa</i> library
JOURNAL	J. Bacteriol., (2002) In press
COMMENT	Contact: Chris K. Raymond Genome Center University of Washington Box 352145, Seattle, WA 98105-2145, USA

516 a	611 c	468 g	159 t
-------	-------	-------	-------

Genome Center  
University of Washington  
Box 352145, Seattle, WA 98105-2145, USA

Query Match 1.6%; Score 23; DB 10; Length 1754;  
 Best Local Similarity 100.0%; Pred. No. 4.6;  
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 884 CGCGCTCGCGCGCTGCGCTGG 906  
 Db 343 CGCGCTCGCGCGCTGCGCTGG 321  
 |||||

RESULT 6  
 AL825430/c  
 LOCUS AL825430 p:335 Triticum aestivum cDNA clone B08\_p335\_plate\_15, mRNA  
 DEFINITION sequence.  
 ACCESSION AL825430  
 VERSION AL825430.1 GI:21836951  
 KEYWORDS EST.  
 SOURCE Triticum aestivum (bread wheat)  
 ORGANISM Triticum aestivum

REFERENCE 1 (bases 1 to 537)  
 AUTHORS Willson, I., Beswick, R., Shepherd, S., Barker, G., Parker, J., Owen, P.,  
 Edwards, D., Coghill, J., Holdsworth, M., Lenton, J., Shewry, P. and  
 Edwards, K.  
 TITLE A BBSRC-funded wheat EST resource for the academic community  
 JOURNAL Unpublished  
 COMMENT Contact: Barker G  
 Institute of Arable Crop Research  
 Long Ashton, Bristol BS41 9AF United Kingdom.

FEATURES  
 source  
 1. .537  
 /organism="Triticum aestivum"  
 /mol\_type="mRNA"  
 /cultivar="mercia"  
 /db\_xref="taxon:4565"  
 /clone="B08\_p335\_plate\_15"  
 /tissue\_type="salt stressed seedlings"  
 /dev\_stage="21 days old"  
 /clone\_lib="P:335"  
 BASE COUNT 86 a 216 c 147 g 87 t 1 others  
 ORIGIN

Query Match 1.5%; Score 22; DB 9; Length 537;  
 Best Local Similarity 100.0%; Pred. No. 9.9;  
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1274 CGCGAGCTGCGCGCGGAGGC 1295  
 Db 117 CGCGAGCTGCGCGCGGAGGC 96  
 |||||

RESULT 7  
 BE586183/c  
 LOCUS BE586183  
 DEFINITION Spike cDNA library Triticum aestivum graminearum infected  
 EST7-4-5\_C02\_7elt7\_006 RSU wheat Fusarium graminearum infected  
 EST7-4-5\_C02\_7elt7\_006, mRNA sequence.  
 ACCESSION BE586183  
 VERSION BE586183.1 GI:9839203  
 KEYWORDS EST.  
 SOURCE Triticum aestivum (bread wheat)  
 ORGANISM Triticum aestivum

REFERENCE 1 (bases 1 to 687)  
 AUTHORS Fellers, J.P., Li, W.L., Hill-Ambroz, K., Matthews, A. and Gill, B.S.  
 TITLE The structure and function of the expressed portion of the wheat  
 genomes - Kansas State University. Fusarium graminearum infected  
 spike cDNA library  
 JOURNAL Unpublished

CONTACT: John Fellers  
 US Department of Agriculture, Agriculture Research Service, Plant  
 Science and Entomology Unit  
 Dept. of Plant Pathology, 4006 Throckmorton Hall, Kansas State  
 University, Manhattan, KS 66506, USA  
 Tel: 785-532-2367  
 Fax: 785-532-6167  
 Email: jpf@alfalfa.ksu.edu  
 Sequence have been trimmed to remove vector sequence and low  
 quality sequence with phred score less than 20  
 Seq primer: T7.

FEATURES  
 source  
 1. .687  
 /organism="Triticum aestivum"  
 /mol\_type="mRNA"  
 /cultivar="Sumai3"  
 /db\_xref="taxon:4565"  
 /clone="Est7-4-5\_C02\_7elt7\_006"  
 /tissue\_type="Spike"  
 /dev\_stage="Adult plant"  
 /lab\_host="E. coli JM109"  
 /clone\_lib="KSU wheat Fusarium graminearum infected spike  
 cDNA library"  
 /note="Vector: pGEM-T easy; Site\_1: SacII; Site\_2: SpeI;  
 Plants were grown in the greenhouse. Spikes were sprayed  
 with Fusarium graminearum (at what stage). Total RNA, and  
 poly(A) RNA were prepared from infected spikes. cDNA was  
 prepared using the SmartPCR cDNA synthesis kit from  
 Clontech. cDNA was cloned into the pGEM-T easy vector  
 from Promega."

BASE COUNT 140 a 246 c 182 g 117 t 2 others  
 ORIGIN

Query Match 1.5%; Score 22; DB 10; Length 687;  
 Best Local Similarity 100.0%; Pred. No. 11;  
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1274 CGCGAGCTGCGCGCGGAGGC 1295  
 Db 428 CGCGAGCTGCGCGCGGAGGC 407  
 |||||

RESULT 8  
 CB061196/c  
 LOCUS CB061196  
 DEFINITION 4011803 BARC-EMBRAPA 338BOV Bos indicus cDNA clone 338BOV\_1C19  
 Unknown, mRNA sequence.  
 ACCESSION CB061196  
 VERSION CB061196.1 GI:27799483  
 KEYWORDS EST.  
 SOURCE Bos indicus (zebu)  
 ORGANISM Bos indicus

REFERENCE 1 (bases 1 to 265)  
 AUTHORS da Mota, A.F., Sonstegard, T.S., Van Tassel, C.P., Matukumalli, L.K.,  
 Wood, D.L., Capuco, A.V., Brito, M.A.P., Martinez, M.L., Connor, E.E.,  
 Machado, M.A. and Coutinho, L.L.  
 TITLE Construction and Characterization of cDNA Libraries Generated from  
 Mammary Gland Tissues of Holstein (Bos taurus) and Gir (Bos indicus  
 ) Cattle  
 JOURNAL Unpublished  
 COMMENT Contact: Adilson F. da Mota  
 Gene Evaluation and Mapping Laboratory  
 USDA, ARS, Animal and Natural Resources Institute  
 Bldg. 200 Rm3 BARC-East, Beltsville, MD 20705, USA  
 Tel: 3015048456  
 Fax: 3015048414  
 Email: amota@npg1.embrapa.br  
 Single pass sequencing. Bases called and trimmed with phred  
 0.000925 using options '-trim\_alt' '-trim\_fasta. Vector identified  
 by cross\_match using options '-minmatch 12 -minscore 12

## PCR Primers

FORWARD: GTTTCAGTCACGAGCTTG  
 BACKWARD: TGAGCGGATCAATTCACACAG

Plate: 1 row: C column: 19

Seq primer: GTTTCAGTCACGAGCTTG

High quality sequence stop: 265.

Location/Qualifiers

## FEATURES

source

1. 265

/organism="Bos indicus"

/mol\_type="mRNA"

/strain="Brazilian Dairy Gir"

/db\_xref="taxon:9915"

/clone="338BOV\_1C19"

/sex="female"

/tissue\_type="teat cistern and alveoli"

/cell\_type="multiple"

/dev\_stage="involted"

/lab\_host="K-12"

/clone\_lib="BARC-EMBRAPA 338BOV"

/note="Organ: mammary; Vector: pUC 118; Site\_1: HincII; Site\_2: HincII; This mammary-derived cDNA library was created as part of a collaborative project between the ARS Gene Evaluation and Mapping Laboratory and the EMBRAPA Dairy Cattle Research Center under the sponsorship of USDA , ARS/EMBRAPA-LABEX program in animal genomics. RNA extracted on 6/8/02, RT with Superscript II at 37 deg C annealing temperature. Cow AM1."

52 a 75 c 92 g 46 t

1.4%; Score 21; DB 14; Length 265;

Best Local Similarity 100.0%; Pred. No. 25;

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 900 GCGCTGGCCATCGCGTGGC 920

Db 255 GCGCTGGCCATCGCGTGGC 235

1. 369

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clone="IMAGE:1579622"

/lab\_host="DH10B"

/clone\_lib="Soares\_NFL\_T\_GBC\_S1"

/note="Organ: pooled; Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site\_1: Not I; Site\_2: Eco RI;

1. 369

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clone="IMAGE:1579622"

/lab\_host="DH10B"

/clone\_lib="Soares\_NFL\_T\_GBC\_S1"

/note="Organ: pooled; Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site\_1: Not I; Site\_2: Eco RI;

1. 369

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clone="IMAGE:1579622"

/lab\_host="DH10B"

/clone\_lib="Soares\_NFL\_T\_GBC\_S1"

/note="Organ: pooled; Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site\_1: Not I; Site\_2: Eco RI;

1. 369

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clone="IMAGE:1579622"

/lab\_host="DH10B"

/clone\_lib="Soares\_NFL\_T\_GBC\_S1"

/note="Organ: pooled; Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site\_1: Not I; Site\_2: Eco RI;

1. 369

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clone="IMAGE:1579622"

/lab\_host="DH10B"

/clone\_lib="Soares\_NFL\_T\_GBC\_S1"

/note="Organ: pooled; Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site\_1: Not I; Site\_2: Eco RI;

1. 369

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clone="IMAGE:1579622"

/lab\_host="DH10B"

/clone\_lib="Soares\_NFL\_T\_GBC\_S1"

/note="Organ: pooled; Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site\_1: Not I; Site\_2: Eco RI;

1. 369

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clone="IMAGE:1579622"

/lab\_host="DH10B"

/clone\_lib="Soares\_NFL\_T\_GBC\_S1"

/note="Organ: pooled; Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site\_1: Not I; Site\_2: Eco RI;

Equal amounts of plasmid DNA from three normalized libraries (fetal lung NBHL19W, testis NHT, and B-cell NCI-CGAP-GCB1) were mixed, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from pools of 5,000 clones made from the same 3 libraries. The pools consisted of I.M.A.G.E. clones 297480-302087, 682632-687239, 726408-728711, and 729096-731399. Subtraction by Bento Soares and M. Fatima Bonaldo.

BASE COUNT 85 a 87 c 107 g 90 t  
 ORIGIN

Query Match 1.4%; Score 21; DB 9; Length 369;

Best Local Similarity 100.0%; Pred. No. 28;

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1268 GCAGCGCCGAGCTGCGCCG 1288

Db 124 GCAGCGCCGAGCTGCGCCG 144

1. 501

/organism="Oryza sativa (indica cultivar-group)"

/mol\_type="mRNA"

/db\_xref="taxon:39946"

/clone="NL\_4\_N22"

/tissue\_type="Entire leaf tissue"

/dev\_stage="35 day-old seedlings"

/clone\_lib="Drought stress (leaf)"

/note="Organ: Leaf; Vector: T7T3Pac; ESTs from normalized leaf cDNA Library from drought stressed seedlings"

52 a 140 c 226 g 82 t 1 others

1. 501

/organism="Oryza sativa (indica cultivar-group)"

/mol\_type="mRNA"

/db\_xref="taxon:39946"

/clone="NL\_4\_N22"

/tissue\_type="Entire leaf tissue"

/dev\_stage="35 day-old seedlings"

/clone\_lib="Drought stress (leaf)"

/note="Organ: Leaf; Vector: T7T3Pac; ESTs from normalized leaf cDNA Library from drought stressed seedlings"

52 a 140 c 226 g 82 t 1 others

1. 501

/organism="Oryza sativa (indica cultivar-group)"

/mol\_type="mRNA"

/db\_xref="taxon:39946"

/clone="NL\_4\_N22"

/tissue\_type="Entire leaf tissue"

/dev\_stage="35 day-old seedlings"

/clone\_lib="Drought stress (leaf)"

/note="Organ: Leaf; Vector: T7T3Pac; ESTs from normalized leaf cDNA Library from drought stressed seedlings"

52 a 140 c 226 g 82 t 1 others

1. 501

/organism="Oryza sativa (indica cultivar-group)"

/mol\_type="mRNA"

/db\_xref="taxon:39946"

/clone="NL\_4\_N22"

/tissue\_type="Entire leaf tissue"

/dev\_stage="35 day-old seedlings"

/clone\_lib="Drought stress (leaf)"

/note="Organ: Leaf; Vector: T7T3Pac; ESTs from normalized leaf cDNA Library from drought stressed seedlings"

52 a 140 c 226 g 82 t 1 others

1. 501

/organism="Oryza sativa (indica cultivar-group)"

/mol\_type="mRNA"

/db\_xref="taxon:39946"

/clone="NL\_4\_N22"

/tissue\_type="Entire leaf tissue"

/dev\_stage="35 day-old seedlings"

/clone\_lib="Drought stress (leaf)"

/note="Organ: Leaf; Vector: T7T3Pac; ESTs from normalized leaf cDNA Library from drought stressed seedlings"

52 a 140 c 226 g 82 t 1 others

1. 501

/organism="Oryza sativa (indica cultivar-group)"

/mol\_type="mRNA"

/db\_xref="taxon:39946"

/clone="NL\_4\_N22"

/tissue\_type="Entire leaf tissue"

/dev\_stage="35 day-old seedlings"

/clone\_lib="Drought stress (leaf)"

/note="Organ: Leaf; Vector: T7T3Pac; ESTs from normalized leaf cDNA Library from drought stressed seedlings"

52 a 140 c 226 g 82 t 1 others

1. 501

/organism="Oryza sativa (indica cultivar-group)"

/mol\_type="mRNA"

/db\_xref="taxon:39946"

/clone="NL\_4\_N22"

/tissue\_type="Entire leaf tissue"

/dev\_stage="35 day-old seedlings"

/clone\_lib="Drought stress (leaf)"

/note="Organ: Leaf; Vector: T7T3Pac; ESTs from normalized leaf cDNA Library from drought stressed seedlings"

52 a 140 c 226 g 82 t 1 others

1. 501

/organism="Oryza sativa (indica cultivar-group)"

/mol\_type="mRNA"

/db\_xref="taxon:39946"

/clone="NL\_4\_N22"

/tissue\_type="Entire leaf tissue"

/dev\_stage="35 day-old seedlings"

/clone\_lib="Drought stress (leaf)"

/note="Organ: Leaf; Vector: T7T3Pac; ESTs from normalized leaf cDNA Library from drought stressed seedlings"

52 a 140 c 226 g 82 t 1 others

1. 501

/organism="Oryza sativa (indica cultivar-group)"

/mol\_type="mRNA"

/db\_xref="taxon:39946"

/clone="NL\_4\_N22"

/tissue\_type="Entire leaf tissue"

/dev\_stage="35 day-old seedlings"

/clone\_lib="Drought stress (leaf)"

/note="Organ: Leaf; Vector: T7T3Pac; ESTs from normalized leaf cDNA Library from drought stressed seedlings"

52 a 140 c 226 g 82 t 1 others

1. 501

/organism="Oryza sativa (indica cultivar-group)"

/mol\_type="mRNA"

/db\_xref="taxon:39946"

/clone="NL\_4\_N22"

/tissue\_type="Entire leaf tissue"

/dev\_stage="35 day-old seedlings"

/clone\_lib="Drought stress (leaf)"

/note="Organ: Leaf; Vector: T7T3Pac; ESTs from normalized leaf cDNA Library from drought stressed seedlings"

52 a 140 c 226 g 82 t 1 others

1. 501

/organism="Oryza sativa (indica cultivar-group)"

/mol\_type="mRNA"

/db\_xref="taxon:39946"

/clone="NL\_4\_N22"

/tissue\_type="Entire leaf tissue"

/dev\_stage="35 day-old seedlings"

/clone\_lib="Drought stress (leaf)"

/note="Organ: Leaf; Vector: T7T3Pac; ESTs from normalized leaf cDNA Library from drought stressed seedlings"

52 a 140 c 226 g 82 t 1 others

1. 501

/organism="Oryza sativa (indica cultivar-group)"

/mol\_type="mRNA"

/db\_xref="taxon:39946"

/clone="NL\_4\_N22"

/tissue\_type="Entire leaf tissue"

/dev\_stage="35 day-old seedlings"

/clone\_lib="Drought stress (leaf)"

/note="Organ: Leaf; Vector: T7T3Pac; ESTs from normalized leaf cDNA Library from drought stressed seedlings"

52 a 140 c 226 g 82 t 1 others

1. 501

/organism="Oryza sativa (indica cultivar-group)"

/mol\_type="mRNA"

/db\_xref="taxon:39946"

/clone="NL\_4\_N22"

/tissue\_type="Entire leaf tissue"

/dev\_stage="35 day-old seedlings"

/clone\_lib="Drought stress (leaf)"

/note="Organ: Leaf; Vector: T7T3Pac; ESTs from normalized leaf cDNA Library from drought stressed seedlings"

52 a 140 c 226 g 82 t 1 others

1. 501

/organism="Oryza sativa (indica cultivar-group)"

/mol\_type="mRNA"

/db\_xref="taxon:39946"

/clone="NL\_4\_N22"

/tissue\_type="Entire leaf tissue"

/dev\_stage="35 day-old seedlings"

/clone\_lib="Drought stress (leaf)"

/note="Organ: Leaf; Vector: T7T3Pac; ESTs from normalized leaf cDNA Library from drought stressed seedlings"

52 a 140 c 226 g 82 t 1 others

1. 501

/organism="Oryza sativa (indica cultivar-group)"

/mol\_type="mRNA"

/db\_xref="taxon:39946"

/clone="NL\_4\_N22"

/tissue\_type="Entire leaf tissue"

/dev\_stage="35 day-old seedlings"

/clone\_lib="Drought stress (leaf)"

/note="Organ: Leaf; Vector: T7T3Pac; ESTs from normalized leaf cDNA Library from drought stressed seedlings"

52 a 140 c 226 g 82 t 1 others

1. 501

/organism="Oryza sativa (indica cultivar-group)"

/mol\_type="mRNA"

/db\_xref="taxon:39946"

/clone="NL\_4\_N22"

/tissue\_type="Entire leaf tissue"

/dev\_stage="35 day-old seedlings"

DNA sequencing by: Washington University Genome Sequencing Center  
 Clone distribution: NCI-CGAP clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL, send email to:  
 infoimage.llnl.gov  
 Seq primer: -400P from Gibco  
 High quality sequence stop: 268.

## FEATURES

## source

1. .286

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clone="IMAGE:3032243"

/lab\_host="DH10B"

/clone\_lib="NCI\_CGAP\_Kid11"

/note="Organ: Kidney; Vector: pT73D-Pac (Pharmacia) with  
 a modified polylinker; Site\_1: Not 1; Site\_2: Eco RI;  
 Plasmid DNA from the normalized library NCI\_CGAP\_Kid3 was  
 prepared, and ss circles were made in vitro. Following HAP  
 purification, this DNA was used as tracer in a subtractive  
 hybridization reaction. The driver was PCR-amplified cDNAs  
 from a pool of 5,000 clones made from the same library  
 (cloneIDs 1322376-1323911, 1456007-1456775, and  
 1500552-1502855). Subtraction by Bento Soares and M.  
 Fatima Bonaudo."

76 a 56 c 71 g 83 t

## BASE COUNT

## ORIGIN

Query Match 1.4%; Score 20; DB 9; Length 286;

Best Local Similarity 100.0%; Pred. No. 80;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1268 GCAGCGCCGCGAGCTGGCCC 1287

Db 147 GCAGCGCCGCGAGCTGGCCC 166

## RESULT 13

## CB042199

## LOCUS

DEFINITION 4007144 BARC-EMBRAPA 316BOV Bos indicus cDNA clone 316BOV\_IG12

Unknown, mRNA sequence.

ACCESSION CB042199

VERSION CB042199.1 GI:27761444

KEYWORDS EST.

SOURCE Bos indicus (zebu)

ORGANISM Bos indicus

REFERENCE 1 (bases 1 to 289)

AUTHORS Wood, D.L., Capuco, A.V., Brito, M.A.P., Martinez, M.L., Connor, E.E.,

Machado, M.A. and Coutinho, L.

Construction and Characterization of cDNA Libraries Generated from

Mammary Gland Tissues of Holstein (Bos taurus) and Gir (Bos indicus)

) Cattle

JOURNAL Unpublished

COMMENT Contact: Adilson F. da Mota

Gene Evaluation and Mapping Laboratory

USDA, ARS, Animal and Natural Resources Institute

Bldg. 200 Rm3 BARC-East, Beltsville, MD 20705, USA

Tel: 3015048456

Fax: 3015048414

Email: amota@cnpgl.embrapa.br

Single pass sequencing. Bases called and trimmed with phred

0.000925 using options -trim\_alt -trim\_fasta. Vector identified

by cross\_match using options -minmatch 12 -minscore 18

PCR Primers

FORWARD: GTTTTCCCGAGTCACGCGTTG

BACKWARD: TGAGCGGATACAAATTCACACAG

Plate: 1 row: G column: 12

Seq primer: GTTTTCCCGAGTCACGCGTTG

High quality sequence stop: 289.

## RESULT 11

## BE041029

## LOCUS

DEFINITION OF16F05 OF Oryza sativa cDNA 5' similar to histone-like dna-binding

protein pf 1, mRNA sequence.

ACCESSION BE041029

VERSION BE041029.1 GI:8336700

KEYWORDS EST.

SOURCE Oryza sativa

ORGANISM Oryza sativa

REFERENCE 1 (bases 1 to 792)

AUTHORS Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;

Ehrhartoideae; Oryzaceae; Oryza.

Bohner, H.J., Borchert, C., Brazille, S., Brooks, J., Eaton, M., Ferrea

, H., Kawasaki, S., McCollough, A., Michalowski, C.B., Palacio, C.,

Scara, G., Wheeler, M. and Zepeda, G.R.

Functional Genomics of Plant Stress Tolerance

Unpublished

CONTACT: Michalowski, C.B.

University of Arizona

Bio Sciences West room 513, Tucson, AZ 85721, USA

Tel: 520-621-7982

Fax: 520-621-1697

Email: cbm@u.arizona.edu

An open reading frame exists.

Location/Qualifiers

1. .792

/organism="Oryza sativa"

/mol\_type="mRNA"

/strain="pokkali"

/db\_xref="taxon:4530"

/tissue\_type="entire plant"

/dev\_stage="2 weeks"

/clone\_lib="OF"

/note="1 week 150mM NaCl"

77 a 189 c 181 g 345 t

## BASE COUNT

## ORIGIN

Query Match 1.4%; Score 21; DB 10; Length 792;

Best Local Similarity 100.0%; Pred. No. 36;

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1050 CTGGCGCGATCGAGGGGCTG 1070

Db 54 CTGGCGCGATCGAGGGGCTG 74

## RESULT 12

## AW771204

## LOCUS

DEFINITION hn60b06.x1 NCI\_CGAP\_Kid11 Homo sapiens cDNA clone IMAGE:3032243 3',

mRNA sequence.

ACCESSION AW771204

VERSION AW771204.1 GI:7703257

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 286)

AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.

National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

Tumor Gene Index

Unpublished

CONTACT: Robert Strausberg, Ph.D.

Email: cgapbs-remail.nih.gov

Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.

Emmert-Buck, M.D., Ph.D.

cDNA Library Preparation: M. Bento Soares, Ph.D.

cDNA Library Arrayed by: Greg Lennon, Ph.D.

FEATURES  
source

Location/Qualifiers  
1. .289  
/organism="Bos indicus"  
/mol\_type="mRNA"  
/strain="Brazilian Dairy Gir"  
/db\_xref="taxon:9915"  
/clone="316BOV\_1G12"  
/sex="female"  
/tissue\_type="mammary"  
/cell\_type="epithelium"  
/dev\_stage="involted"  
/lab\_host="K-12"  
/clone\_lib="BARC-EMBRAPA 316BOV"  
/note="Organ: mammary; Vector: pUC19; Site\_1: SmaI; Site\_2: SmaI; This mammary-derived cDNA library was created as part of a collaborative project between the ARS Gene Evaluation and Mapping Laboratory and the EMBRAPA Dairy Cattle Research Center under the sponsorship of USDA , ARS/EMBRAPA-LABEX program in animal genomics. RNA extracted on 6/8/02, RT with Superscript II at 37 deg C, annealing temperature, PCR with 16mer AM5."

BASE COUNT 47 a 100 c 87 g 55 t  
ORIGIN  
Query Match 1.4%; Score 20; DB 14; Length 289;  
Best Local Similarity 100.0%; Pred. No. 81;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 901 CGCTGGGCCATCGCGTCGCG 920  
|||||  
Db 1 CGCTGGGCCATCGCGTCGCG 20

## RESULT 14

BF657011/c 301 bp mRNA linear EST 20-DEC-2000  
LOCUS OV2\_18.A07.g1.A002 Ovary 2 (OV2) Sorghum bicolor cDNA, mRNA  
DEFINITION sequence.

ACCESSION BF657011  
VERSION BF657011.1 GI:11922145  
KEYWORDS EST.  
SOURCE Sorghum bicolor (sorghum)  
ORGANISM Sorghum bicolor

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Sorghum.

1 (bases 1 to 301)  
REFERENCE Cordonnier-Pratt,M.-M., Gingle,A., Marsala,C., Sudman,M. and Pratt L.H.

TITLE An EST database from Sorghum: ovaries of varying immature stages

UNPUBLISHED  
Contact: Cordonnier-Pratt MM  
Laboratory for Genomics and Bioinformatics  
The University of Georgia, Department of Plant Biology  
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA  
Tel: 706 542 1860  
Fax: 706 583 0210  
Email: mmpratt@uga.edu

Seq primer: T7

High quality sequence start: 6

High quality sequence stop: 222

POLYA-Yes.

FEATURES  
source

Location/Qualifiers  
1. .301  
/organism="Sorghum bicolor"  
/mol\_type="mRNA"  
/db\_xref="taxon:4558"  
/clone\_lib="Ovary 2 (OV2)"  
/note="Organ: Mix of ovaries of varying immature stages from 8-week-old plants; Vector: pBluescript II from Lambda Zap II; Site\_1: XhoI; Site\_2: EcoRI; The library was made from poly-A RNA in the cloning vector lambda Zap II. Clones to be sequenced were prepared by mass excision."

BASE COUNT  
ORIGIN

76 a 62 c 83 g 80 t

Query Match 1.4%; Score 20; DB 10; Length 301;  
Best Local Similarity 100.0%; Pred. No. 82;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1305 CTGTTGTGCGCGCTCGC 1324  
|||||

Db 37 CTGTTGTGCGCGCTCGC 18

## RESULT 15

AW139275 318 bp mRNA linear EST 30-OCT-1999  
LOCUS UI-H-B11-aeg-g-12-0-UI.s1 NCI\_CGAP\_Sub3 Homo sapiens cDNA clone  
DEFINITION IMAGE:2719318 3', mRNA sequence.

ACCESSION AW139275

VERSION AW139275.1 GI:6143993

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 318)

AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.

TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

Tumor Gene Index

UNPUBLISHED

Contact: Robert Strausberg, Ph.D.

Email: cgapbs-r@mail.nih.gov

Oligo-dT track not found, Not I site shown in beginning of sequence is likely internal to the message. cDNA Library Preparation: M.B.

Soares Lab Clone distribution: NCI-CGAP clone distribution

Information can be found through the I.M.A.G.E. Consortium/LLNL at:

www-bio.llnl.gov/bbrp/image/image.html

Seq primer: M13 Forward

POLYA-No.

FEATURES  
source

Location/Qualifiers  
1. .318  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:2719318"  
/lab\_host="DH10B (Life Technologies)"  
/clone\_lib="NCI\_CGAP\_Sub3"  
/note="Vector: pT7D-Pac (Pharmacia) with a modified

polylinker; Site\_1: Not I; Site\_2: Eco RI; The NCI\_CGAP\_Sub3 library is a subtracted library derived from the NCI\_CGAP\_Sub1 library, which is a subtracted library derived from BI. BI constitutes a mixture of 21 normalized or subtracted NCI\_CGAP libraries: NCI\_CGAP\_Co4 , NCI\_CGAP\_Pr22, NCI\_CGAP\_Pr28, NCI\_CGAP\_Co10, NCI\_CGAP\_Co16, NCI\_CGAP\_Kid5, NCI\_CGAP\_Kid12, NCI\_CGAP\_Kid3, NCI\_CGAP\_Kid11, NCI\_CGAP\_Lym2, NCI\_CGAP\_Br2, NCI\_CGAP\_Co8, NCI\_CGAP\_C111, NCI\_CGAP\_Lei2, NCI\_CGAP\_Brn23, NCI\_CGAP\_Lu5, NCI\_CGAP\_Lu24, NCI\_CGAP\_Lu19, NCI\_CGAP\_GC4, NCI\_CGAP\_GC6, NCI\_CGAP\_Brn25. These 21 libraries were pooled and a

single-stranded DNA preparation of the resulting mixture was used as a tracer in a subtractive hybridization with

a driver whose composition is detailed below:

NCI\_CGAP\_Kid3 pool 1 LLAM 3334-3337, 3682-3683, 3798-3803 (IMAGE Clones 132376-132391, 1456008-1456775 , 1500552-1502855); NCI\_CGAP\_Kid5 pool 1, LLAM 3338-3342 , 3722-3725, 3776-3778 (IMAGE Clones 1323912-1325831, 1471368-1472903, 1492104-1493255); NCI\_CGAP\_Lu5 pool 1 LLAM 3575-3582, 3851-3854 (IMAGE Clones 1414920-1417991, 1520904-1522439); NCI\_CGAP\_GC4 pool 1 LLAM 3164-3167, 3716-3720, 3733-3735 (IMAGE Clones 1257096-1258631, 1469064-1470983, 1475592-1476743); NCI\_CGAP\_Pr22 pool 1 LLAM 2457-2459, 2758-2759, 3062-3068 (IMAGE Clones 985608-986759, 1101192-1101959, 1217928-1220615);

NCI\_CGAP\_Col0 pool 1 LLAM 2644-2653, 2871-2872 (IMAGE  
Clones 1057416-1061255, 1144584-1145351). Subtraction  
was performed as previously described (Bonaldi, Lennon &  
Soares (1996)): Normalization and Subtraction: Two  
Approaches To Facilitate Gene Discovery. Genome Research  
6, 791-806.

TAG\_LIB=NCI\_CGAP\_GC4  
TAG\_TISSUE=germ cell  
TAG\_SEQ=AAATC\*

BASE COUNT	79 a	70 c	80 g	88 t	1 others
------------	------	------	------	------	----------

Query Match 1.4%; Score 20; DB 9; Length 318;  
Best Local Similarity 100.0%; Pred. No. 83;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1268	GCAGGCCGCGAGCTGGCCC	1287
Db	174	GCAGGCCGCGAGCTGGCCC	193

Search completed: August 14, 2003, 07:04:13  
Job time : 2903 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 14, 2003, 01:55:34 ; Search time 422 Seconds  
(without alignments)  
9275.320 Million cell updates/sec

Title: US-09-701-229-1  
Perfect score: 1450  
Sequence: 1 cgtgctgacggcctcgcca.....tgttgacggcgacgcgcac 1450

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 2552756 seqs, 1349719017 residues

Word size : 0  
Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :				N_Geneseq_19Jun03.*	
1:	/SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*	1:	/SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*	1:	/SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*
2:	/SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*	2:	/SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*	2:	/SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
3:	/SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*	3:	/SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*	3:	/SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*
4:	/SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*	4:	/SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*	4:	/SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
5:	/SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*	5:	/SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*	5:	/SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*
6:	/SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*	6:	/SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*	6:	/SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*
7:	/SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*	7:	/SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*	7:	/SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*
8:	/SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA2001.DAT.*	8:	/SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*	8:	/SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA2003.DAT.*
9:	/SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA2004.DAT.*	9:	/SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA2005.DAT.*	9:	/SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA2006.DAT.*
10:	/SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA2007.DAT.*	10:	/SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA2008.DAT.*	10:	/SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA2009.DAT.*
11:	/SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA2010.DAT.*	11:	/SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA2011.DAT.*	11:	/SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA2012.DAT.*
12:	/SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA2013.DAT.*	12:	/SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA2014.DAT.*	12:	/SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA2015.DAT.*
13:	/SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA2016.DAT.*	13:	/SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA2017.DAT.*	13:	/SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA2018.DAT.*
14:	/SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA2019.DAT.*	14:	/SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA2020.DAT.*	14:	/SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA2021.DAT.*
15:	/SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA2022.DAT.*	15:	/SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA2023.DAT.*	15:	/SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA2024.DAT.*
16:	/SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA2025.DAT.*	16:	/SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA2026.DAT.*	16:	/SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA2027.DAT.*
17:	/SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA2028.DAT.*	17:	/SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA2029.DAT.*	17:	/SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA2030.DAT.*
18:	/SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA2031.DAT.*	18:	/SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA2032.DAT.*	18:	/SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA2033.DAT.*
19:	/SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA2034.DAT.*	19:	/SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA2035.DAT.*	19:	/SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA2036.DAT.*
20:	/SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA2037.DAT.*	20:	/SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA2038.DAT.*	20:	/SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA2039.DAT.*
21:	/SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA2040.DAT.*	21:	/SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA2041.DAT.*	21:	/SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA2042.DAT.*
22:	/SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA2043.DAT.*	22:	/SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA2044.DAT.*	22:	/SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA2045.DAT.*
23:	/SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA2046.DAT.*	23:	/SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA2047.DAT.*	23:	/SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA2048.DAT.*
24:	/SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA2049.DAT.*	24:	/SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA2050.DAT.*	24:	/SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA2051.DAT.*
25:	/SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA2052.DAT.*	25:	/SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA2053.DAT.*	25:	/SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA2054.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	1450	100.0	1450	21	AZ47131 Pseudomonas aerugi
2	1347	92.9	1347	23	AAS54275 Pseudomonas aerugi
3	54	3.7	1200	23	AAS51550 Pseudomonas aerugi
4	36	2.5	116	23	AAS48736 Pseudomonas aerugi
5	36	2.5	1083	22	AAF86639 Pseudomonas aerugi
6	36	2.5	1083	23	AAS44276 Pseudomonas aerugi
7	24	1.7	1080	23	AAS89258 DNA encoding novel
8	24	1.7	1080	23	AAS89801 DNA encoding novel

9	24	1.7	1080	23	AAS90166 DNA encoding novel
10	24	1.7	1317	22	AAS81460 Escherichia coli p
11	24	1.7	1317	23	AAS52295 E. coli DNA for ce
12	24	1.7	2058	23	AAS73012 DNA encoding novel
13	24	1.7	3117	23	AAS69095 DNA encoding novel
14	24	1.7	3117	23	AAS89240 DNA encoding novel
15	24	1.7	4831	23	AAS94542 DNA encoding novel
16	24	1.7	4942	23	AAS89816 DNA encoding novel
17	24	1.7	4944	23	AAS89273 DNA encoding novel
18	24	1.7	4944	23	AAS90181 DNA encoding novel
c 19	22	1.5	27	22	AAF86641 Pseudomonas aerugi
c 20	20	1.4	507	21	AAF07985 Fusarium venenatum
c 21	20	1.4	3511	22	AAI59407 Human polynucleoti
c 22	20	1.4	3511	22	AAI59407 Human c-ski oncopr
c 23	20	1.4	42000	21	AAI63349 Streptomyces globi
c 24	20	1.4	63164	21	AAI63348 Streptomyces globi
c 25	20	1.4	109519	22	AAI08693 Micromonospora DNA
c 26	19	1.3	359	24	ABN23890 Human ORFX polynuc
27	19	1.3	522	20	AAI19400 M. tuberculosis re
28	19	1.3	522	20	AAI19188 M. tuberculosis re
29	19	1.3	966	23	AAS56152 Salmonella typhi D
30	19	1.3	1362	20	AAV59101 S. aureofaciens te
31	19	1.3	1857	24	ABQ90059 M. capsulatus gene
c 32	19	1.3	2169	22	AAI58060 Human polynucleoti
33	19	1.3	2932	13	AAQ25388 TXA2 receptor gene
34	19	1.3	2932	20	AAZ32161 Human thromboxane
35	19	1.3	2932	20	AAZ32162 Human endothelial
36	19	1.3	2932	23	AAS65879 DNA encoding novel
37	19	1.3	2932	24	ABK83654 Human cDNA differe
38	19	1.3	2932	25	ACA56821 Human signalling p
39	19	1.3	2932	25	ABZ42829 Human thromboxane
40	19	1.3	3489	23	AAS4094 Pseudomonas aerugi
41	19	1.3	6390	22	AAF81370 Quorum sensing con
c 42	19	1.3	9333	22	AAK74093 Human immune/haema
c 43	19	1.3	14272	21	AAK75096 Nucleotide sequenc
c 44	19	1.3	21500	23	AAS99633 Propionibacterium
45	19	1.3	27541	22	AAI17185 Streptomyces nous

ALIGNMENTS

RESULT 1  
AAZ47131  
ID AAZ47131 standard; DNA; 1450 BP.

XX AAZ47131;

XX 28-MAR-2000 (first entry)

XX Pseudomonas aeruginosa murD gene.

XX MurD; bacterial cell wall; biosynthesis; inhibitor; D-glutamate;  
XX UDP-N-acetylmuramyl-L-alanine precursor; ds.

XX Pseudomonas aeruginosa.

XX WO9961050-A1.

XX 02-DEC-1999.

XX 26-MAY-1999; 99WO-US11585.

XX 29-MAY-1998; 98US-0087308.

XX (MERI ) MERCK & CO INC.

XX El-Sherbeini M, Azzolina B;

XX WPT; 2000-072548/06.

XX P-PSDB; AAY56047.

XX New nucleic acid encoding the MurD protein of Pseudomonas aeruginosa,

PT used to identify specific inhibitors -

Claim 4; Page 8-9; 35pp; English.

CC This sequence represents the MurD gene from *Pseudomonas aeruginosa*.  
CC MurD is an enzyme involved in biosynthesis of bacterial cell walls,  
CC so inhibiting it inhibits bacterial growth (by preventing  
CC MurD-catalyzed addition of D-glutamate to an alanine residue in the  
CC UDP-N-acetylmuramyl-L-alanine precursor). Cells that express murD  
CC protein can be used: (a) to identify inhibitors of MurD, and (b) for  
CC recombinant production of MurD, which is used to raise specific  
CC antibodies (potential therapeutic inhibitors). Fragments of the murD  
CC gene can be used as probes and primers to detect MurD-encoding sequences  
CC (to identify polymorphic murD genes or to detect *P. aeruginosa*), also  
CC as antisense modulators. The murD inhibitors are useful for treating or  
CC preventing bacterial infections, including those caused by *P. aeruginosa*  
XX Sequence 1450 BP; 227 A; 487 C; 487 G; 249 T; 0 other;  
SQ

SQ Sequence 1450 BP; 227 A; 487 C; 487 G; 249 T; 0 other;

```
Query Match      100.0%; Score 1450; DB 21; Length 1450;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1450: Conservative 0; Mismatches 0; Indels 0;
```

QY	1	CGTGTGATCGCCCTCGCCACCTTGAAGCTCGGTTGAGGACGAAGAGAGCATGAGCCTGA	60
Db	1		
QY	61	TCGCCTCGGACCACTTCGCGATCGTTGTCGGCCTCGCAAGAGCGGCATGCTCCCTGGTGC	120
Db	61		
QY	121	GCTACTGGCGCGCGCGGCTTGCCCTTTCGCCGTGGTCGATACCCGAGAGAACCCCCGG	180
Db	121		
QY	181	AGCTGGCCACCTCGGTCGCCAGTATCCGCAAGTGGAAAGTGCCTTGGCGGCACTCGACG	240
Db	181		
QY	241	CCGAGTTCCTCTGCTCCCGCCGCAACTCTATGTACGCCCCCGGCTTGTCTGCTCGGCACCC	300
Db	241		
QY	301	CTGGCCTGGTACAGCGCCGCGGAAAGCGTGCCTATCTCCGGTGACATCGATCTCTTCG	360
Db	301		
QY	361	CCGCGGAGCGAAAGGCCCGCATCTGTCGCCATCACCGGTTCCAAAGCGAAGACGACCGTGA	420
Db	361		
QY	421	CCACCTTGGTGGCGGAATGGCGGTGGCCGCGGACAAGCGTGTGCGCGTGGCGGGAACCC	480
Db	421		
QY	481	TCGGACCCCGCGCTCGACCTGCTGGCCGAGACATCGAGCTGTAGTGTTGGAGCTGT	540
Db	481		
QY	541	CGAGCTTCCAGCTGGAAACCTCGCATCGCTTCAACGCCGAGGTGGCGACCGTCTGAACG	600
Db	541		
QY	601	TCAGCGAAGACCATATGGATCGCTACGACGCATGGCTGACTACCACTGGCCCAAGACCC	660
Db	601		
QY	661	GGATCTTCGCGGTCCCGCCAGGTCGTGGTGAATCGGCCGATGCCCTGACCCGACCCG	720
Db	661		
QY	721	TGATCCCGATACCGTGCCTGCTGGTTCGGCCTGGAACAAGCGGACTTCAAGGCTT	780
Db	721		

23-MAY-2000; 2000US-206848P.  
26-MAY-2000; 2000US-207727P.  
23-OCT-2000; 2000US-242578P.  
27-NOV-2000; 2000US-253625P.  
22-DEC-2000; 2000US-257931P.  
16-FEB-2001; 2001US-269308P.  
(ELIT-) ELITRA PHARM INC.  
Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;  
Yamamoto RT, Xu HH;  
WPI; 2001-611495/70.  
P-PSDB; AAU36416.  
New polynucleotides for the identification and development of  
antibiotics, comprise sequences of antisense nucleic acids -  
Claim 27; Seq ID No 7912; 51pp; English.  
The invention relates to antisense inhibitors of genes essential to  
prokaryotic cellular proliferation, their use in identifying the  
genes, their use in the discovery of novel antibiotics, the essential  
genes themselves and the encoded proteins. The prokaryotes used are  
Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella  
pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The  
invention is also useful for the identification of potential new targets  
for antibiotic development. The antisense nucleic acids can also be used  
to identify proteins used in proliferation, to express these proteins,  
and to obtain antibodies capable of binding to the expressed proteins.  
The proteins can be used to screen compounds in rational drug discovery  
programmes. The antisense nucleic acid sequence is also useful to screen  
for homologous nucleic acids which are required for cell proliferation in  
a wide variety of organisms. The present sequence encodes an  
essential prokaryotic cellular proliferation protein.  
Note: The sequence data for this patent did not form part  
of the printed specification, but was obtained in electronic  
format directly from WIPO at  
ftp.wipo.int/pub/published\_pct\_sequences.  
Seq  
Sequence 1347 BP; 214 A; 453 C; 227 T; 0 other;

Query Match 92.9%; Score 1347; DB 23; Length 1347;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1347; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

51 ATGAGCCCTGATCGCTCGACACCTTCGCGCATCGTTGCGCCTCGGCAAGAGCGGCATG 110  
1 ATGAGCCCTGATCGCTCGACACCTTCGCGCATCGTTGCGCCTCGGCAAGAGCGGCATG 60  
111 TCCCTGGTGGCGTACCTCGCGCGCGCGGCTTGCCTTTTCGCGGTGCGATACCGGAG 170  
61 TCCCTGGTGGCGTACCTCGCGCGCGCGGCTTGCCTTTTCGCGGTGCGATACCGGAG 120  
171 AACCGCGCGGAGCTGGCCACCTTGGTCCGACAGTATCCGAGGTGGAGTGGTGGCGC 230  
121 AACCGCGCGGAGCTGGCCACCTTGGTCCGACAGTATCCGAGGTGGAGTGGTGGCGC 180  
231 GAACCTGACGCGGAGTCTCTCTGCTCCGCGCGCAACTCTATGTCAGCCCGCGGCTTTCG 290  
181 GAACCTGACGCGGAGTCTCTCTGCTCCGCGCGCAACTCTATGTCAGCCCGCGGCTTTCG 240  
291 CTGCGCACCCCTGCGCTGGTATACAGCCCGCGCGAAGGCGTGGCGATCTCCGGTGACATC 350  
241 CTGCGCACCCCTGCGCTGGTATACAGCCCGCGCGAAGGCGTGGCGATCTCCGGTGACATC 300  
351 GATCTCTTGGCCGCGGAGCGAAGGCCCGCGATTCGTCGCATCACCGGTTTCCAAACCGAAG 410  
301 GATCTCTTGGCCGCGGAGCGAAGGCCCGCGATTCGTCGCATCACCGGTTTCCAAACCGAAG 360  
411 AGCACCGTGACACCCCTGTTGGCGCAAAATGGCGGTGGCGGACAAAGCGTGTCCCGCTC 470  
361 AGCACCGTGACACCCCTGTTGGCGCAAAATGGCGGTGGCGGACAAAGCGTGTCCCGCTC 420

QY 471 GCGGCAACCTCGGCACCCCGCGCTCGACCTGCTGGCCGAGACATCGAGCTGTACGTC 530  
DB 421 GCGGCAACCTCGGCACCCCGCGCTCGACCTGCTGGCCGAGACATCGAGCTGTACGTC 480  
QY 531 TTGGAGCTGTGAGCTTCCAGCTGGAACCTGCGATCGCATCGCCCTCAACCGCGAGGTGCGGACC 590  
DB 481 TTGGAGCTGTGAGCTTCCAGCTGGAACCTGCGATCGCATCGCCCTCAACCGCGAGGTGCGGACC 540  
QY 591 GTGCTGAACCTCAGCGAAGACCATATGATGCTGCTAGCAGCGCATGGCTGACTACACCTG 650  
DB 541 GTGCTGAACCTCAGCGAAGACCATATGATGCTGCTAGCAGCGCATGGCTGACTACACCTG 600  
QY 651 GCCAAGCACCGGATCTTCCGCGGTCCCGCGAGGTGCTGGAATCGCGCCGATGCCCTG 710  
DB 601 GCCAAGCACCGGATCTTCCGCGGTCCCGCGAGGTGCTGGAATCGCGCCGATGCCCTG 660  
QY 711 ACCGACCGCTGATGCGCGGATACCGTGGCTGCTGGTGGCTGGAACGACCGGAC 770  
DB 661 ACCGACCGCTGATGCGCGGATACCGTGGCTGCTGGTGGCTGGAACGACCGGAC 720  
QY 771 TTCAGGCTTTGGGCTGATCGAGGAAGACGCCAAGAGTGGCTGGCGTTCAGTTCGAC 830  
DB 721 TTCAGGCTTTGGGCTGATCGAGGAAGACGCCAAGAGTGGCTGGCGTTCAGTTCGAC 780  
QY 831 AAGCTGCTCCGCGTTGGCGAACTGAAGATCCGTGGCGGCCCAACAATATTTCAACCGGCTC 890  
DB 781 AAGCTGCTCCGCGTTGGCGAACTGAAGATCCGTGGCGGCCCAACAATATTTCAACCGGCTC 840  
QY 891 GCGCGCTGCGGCTGGGCGCATGCGGTGCGGCTGCGGTTTCGAGCCGATCGTGGCGGCTG 950  
DB 841 GCGCGCTGCGGCTGGGCGCATGCGGTGCGGCTGCGGTTTCGAGCCGATCGTGGCGGCTG 900  
QY 951 AAGCGGTTTCCGCGCTGGCTCATCGCTGCCAGTGGGTACCGAGCGCAGCGGCTGAGC 1010  
DB 901 AAGCGGTTTCCGCGCTGGCTCATCGCTGCCAGTGGGTACCGAGCGCAGCGGCTGAGC 960  
QY 1011 TACTACGACGATTCCAAGGCCACCAACGTCGCGCGCGCCCTGGCGCGGATCGAGGGGCTG 1070  
DB 961 TACTACGACGATTCCAAGGCCACCAACGTCGCGCGCGCCCTGGCGCGGATCGAGGGGCTG 1020  
QY 1071 GGTGCCGACATCGAGCGCAAGCTGTGCTGCTGCGCGCGGAGAGCGGAAGGGCCCGAT 1130  
DB 1021 GGTGCCGACATCGAGCGCAAGCTGTGCTGCTGCGCGCGGAGAGCGGAAGGGCCCGAT 1080  
QY 1131 TTCCATGACCTGCGCGAGCGGTGCGCGCTTCTCCCGGGGGTGTACTGCTTGGCGCT 1190  
DB 1081 TTCCATGACCTGCGCGAGCGGTGCGCGCTTCTCCCGGGGGTGTACTGCTTGGCGCT 1140  
QY 1191 GACGCGGGCTGATTGGCCAGGCACTGGGCAACGCGGTACCGCTGGTGGCGTTCGCAACG 1250  
DB 1141 GACGCGGGCTGATTGGCCAGGCACTGGGCAACGCGGTACCGCTGGTGGCGTTCGCAACG 1200  
QY 1251 CTGACGAAAGCAGTCCGCGAGCGCCCGAGCTGGCGCGGAGGCGATGCGGTGCTGTC 1310  
DB 1201 CTGACGAAAGCAGTCCGCGAGCGCCCGAGCTGGCGCGGAGGCGATGCGGTGCTGTC 1260  
QY 1311 TCGCGCGCTGCGCGAGCGCTGGACATGTTCAAGAACTTCGAAGAACGCGGAGCGCTGTC 1370  
DB 1261 TCGCGCGCTGCGCGAGCGCTGGACATGTTCAAGAACTTCGAAGAACGCGGAGCGCTGTC 1320  
QY 1371 GCCAAAGCGGTAGAGGAGCTAGCGTGA 1397  
DB 1321 GCCAAAGCGGTAGAGGAGCTAGCGTGA 1347

RESULT 3  
AAS1550  
ID AAS1550 standard; DNA; 1200 BP.  
XX  
AC AAS1550;  
XX 13-FEB-2002 (first entry)  
DT

```

XX DE Pseudomonas aeruginosa DNA for cellular proliferation protein #135.
XX DE
XX DE
KW Antisense; ds; prokaryotic cellular proliferation gene;
KW Antibiotic; antibacterial; drug design.
XX DE
XX OS Pseudomonas aeruginosa.
XX OS
XX PN WO200170955-A2.
XX PN
XX PD 27-SEP-2001.
XX PD
XX PF 21-MAR-2001; 2001WO-US09180.
XX PF
XX PR 21-MAR-2000; 2000US-191078P.
XX PR
XX PR 23-MAY-2000; 2000US-206848P.
XX PR
XX PR 26-MAY-2000; 2000US-207727P.
XX PR
XX PR 23-OCT-2000; 2000US-242578P.
XX PR
XX PR 27-NOV-2000; 2000US-253625P.
XX PR
XX PR 22-DEC-2000; 2000US-257931P.
XX PR
XX PR 16-FEB-2001; 2001US-269308P.
XX PR
XX PA (ELIT-) ELITRA PHARM INC.
XX PA
XX PI Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;
XX PI Yamamoto RT, Xu HH;
XX PI
XX DR WPI; 2001-611495/70.
XX DR
XX DR P-PSDB; AAU33691.
XX DR
XX PT New polynucleotides for the identification and development of
XX PT antibiotics, comprise sequences of antisense nucleic acids -
XX PT
XX PS Claim 27; Seq ID No 4132; 511pp; English.
XX PS
XX CC The invention relates to antisense inhibitors of genes essential to
XX CC prokaryotic cellular proliferation, their use in identifying the
XX CC genes, their use in the discovery of novel antibiotics, the essential
XX CC genes themselves and the encoded proteins. The prokaryotes used are
XX CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
XX CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
XX CC invention is also useful for the identification of potential new targets
XX CC for antibiotic development. The antisense nucleic acids can also be used
XX CC to identify proteins used in proliferation, to express these proteins,
XX CC and to obtain antibodies capable of binding to the expressed proteins.
XX CC The proteins can be used to screen compounds in rational drug discovery
XX CC programmes. The antisense nucleic acid sequence is also useful to screen
XX CC for homologous nucleic acids which are required for cell proliferation in
XX CC a wide variety of organisms. The present sequence encodes an
XX CC essential prokaryotic cellular proliferation protein.
XX CC Note: The sequence data for this patent did not form part
XX CC of the printed specification, but was obtained in electronic
XX CC format directly from WIPO at
XX CC ftp.wipo.int/pub/published_pct_sequences.
XX CC
XX SQ Sequence 1200 BP; 157 A; 384 C; 402 G; 257 T; 0 other;

Query Match 3.7%; Score 54; DB 23; Length 1200;
Best Local Similarity 100.0%; Pred. No. 2e-14;
Matches 54; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1397 ATGCTGCTGCTGTTGGCCCTTCCGCTGCTGTTGAGCGGCGACGGCATC 1450
|||||
DB 1 ATGCTGCTGCTGTTGGCCCTTCCGCTGCTGTTGAGCGGCGACGGCATC 54

RESULT 4
AAS48736/C
ID AAS48736 standard; DNA; 116 BP.
XX
XX AAS48736;
XX
XX 13-FEB-2002 (first entry)
XX

```

```

XX DE Pseudomonas aeruginosa cellular proliferation inhibitory sequence #222.
XX DE
XX DE
KW Antisense; ss; prokaryotic cellular proliferation;
KW Antibiotic; antibacterial; drug design.
XX DE
XX OS Pseudomonas aeruginosa.
XX OS
XX PN WO200170955-A2.
XX PN
XX PD 27-SEP-2001.
XX PD
XX PF 21-MAR-2001; 2001WO-US09180.
XX PF
XX PR 21-MAR-2000; 2000US-191078P.
XX PR
XX PR 23-MAY-2000; 2000US-206848P.
XX PR
XX PR 26-MAY-2000; 2000US-207727P.
XX PR
XX PR 23-OCT-2000; 2000US-242578P.
XX PR
XX PR 27-NOV-2000; 2000US-253625P.
XX PR
XX PR 22-DEC-2000; 2000US-257931P.
XX PR
XX PR 16-FEB-2001; 2001US-269308P.
XX PR
XX PA (ELIT-) ELITRA PHARM INC.
XX PA
XX PI Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;
XX PI Yamamoto RT, Xu HH;
XX PI
XX DR WPI; 2001-611495/70.
XX DR
XX DR New polynucleotides for the identification and development of
XX DR antibiotics, comprise sequences of antisense nucleic acids -
XX DR
XX PS Claim 1; Seq ID No 1313; 511pp; English.
XX PS
XX CC The invention relates to antisense inhibitors of genes essential to
XX CC prokaryotic cellular proliferation, their use in identifying the
XX CC genes, their use in the discovery of novel antibiotics, the essential
XX CC genes themselves and the encoded proteins. The prokaryotes used are
XX CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
XX CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
XX CC invention is also useful for the identification of potential new targets
XX CC for antibiotic development. The antisense nucleic acids can also be used
XX CC to identify proteins used in proliferation, to express these proteins,
XX CC and to obtain antibodies capable of binding to the expressed proteins.
XX CC The proteins can be used to screen compounds in rational drug discovery
XX CC programmes. The antisense nucleic acid sequence is also useful to screen
XX CC for homologous nucleic acids which are required for cell proliferation in
XX CC a wide variety of organisms. The present sequence is an antisense
XX CC oligonucleotide of the invention.
XX CC Note: The sequence data for this patent did not form part
XX CC of the printed specification, but was obtained in electronic
XX CC format directly from WIPO at
XX CC ftp.wipo.int/pub/published_pct_sequences.
XX CC
XX SQ Sequence 116 BP; 27 A; 37 C; 44 G; 8 T; 0 other;

Query Match 2.5%; Score 36; DB 23; Length 116;
Best Local Similarity 100.0%; Pred. No. 3.3e-06;
Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1415 CCCTTCCCGTCGCGCTGTTGAGCGGCGACGGCATC 1450
|||||
DB 116 CCCTTCCCGTCGCGCTGTTGAGCGGCGACGGCATC 81

RESULT 5
AAF86639
ID AAF86639 standard; DNA; 1083 BP.
XX
XX AAF86639;
XX
XX 17-JUL-2001 (first entry)
XX

```

DE Pseudomonas aeruginosa mray gene.  
XX  
KW Mray: phospho-N-acetylmuramoyl-pentapeptide translocase;  
KW peptidoglycan biosynthesis; Bacterial cell wall; infection;  
KW drug screening; antibacterial; Pseudomonas; G+C rich bacterium; ds.  
XX  
OS Pseudomonas aeruginosa.  
XX  
FH Key Location/Qualifiers  
FT CDS 1..1083  
FT /\*tag= a  
FT /product= "Mray protein"  
FT /function= "Catalyses the transfer of N-acetylmuramic  
FT acid peptide to a bactoprenol phosphate  
FT carrier molecule in peptidoglycan  
FT biosynthesis"  
XX  
PN WO200125251-A1.  
XX  
XX 12-APR-2001.  
XX  
XX 29-SEP-2000; 2000WO-US27056.  
XX  
XX 04-OCT-1999; 99US-0157580.  
XX  
XX (MERI ) MERCK & CO INC.  
XX  
XX El-Sherbeini M, Azzolina B;  
XX  
XX WPI: 2001-308221/32.  
XX P-PSDB; AAB73487.  
XX  
XX New Mray gene and enzyme of Pseudomonas aeruginosa, useful in vitro  
PT assays for screening antibacterial compounds that target cell wall  
PT biosynthesis, particularly for screening antibiotics against  
PT Pseudomonas  
XX  
XX Claim 2; Fig 1; 22pp; English.  
XX  
XX This sequence represents the Pseudomonas aeruginosa mray gene, which  
CC encodes a phospho-N-acetylmuramoyl-pentapeptide translocase designated  
CC Mray. The Mray protein is involved in bacterial cell wall biosynthesis.  
CC It catalyses the first step of the membrane cycle of peptidoglycan  
CC biosynthesis, i.e., the transfer of an N-acetylmuramic acid peptide to a  
CC bactoprenol phosphate carrier molecule. The Mray protein is useful in  
CC in vitro assays to screen for antibacterial compounds that target cell  
CC wall biosynthesis. Inhibitors of the Mray protein are useful in  
CC preventing the growth of Pseudomonas and other G+C rich bacteria.  
CC Pseudomonas aeruginosa is an opportunistic pathogen which causes  
CC infections in patients with burns, neutropenia, or cystic fibrosis.  
CC Primers or probes derived from the mray gene are useful in nucleic acid  
CC amplification-based assays for detecting the presence of a polynucleotide  
CC encoding Pseudomonas aeruginosa Mray protein.  
XX  
XX Sequence 1083 BP; 164 A; 350 C; 333 G; 236 T; 0 other;  
SQ  
Query Match 2.5%; Score 36; DB 22; Length 1083;  
Best Local Similarity 100.0%; Pred. No. 2.6e-06;  
Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 2 GTGCTGATGGCGCTCGCCACCTTGAAGTCGCTTGA 37  
|||||  
Db 1048 GTGCTGATGGCGCTCGCCACCTTGAAGTCGCTTGA 1083  
RESULT 6  
AAS54276  
ID AAS54276 standard; DNA; 1083 BP.  
XX  
AC AAS54276;  
XX  
XX 13-FEB-2002 (first entry)  
XX  
XX

DE Pseudomonas aeruginosa DNA for cellular proliferation protein #407.  
XX  
KW Antisense; ds; prokaryotic cellular proliferation gene;  
KW antibiotic; antibacterial; drug design.  
XX  
OS Pseudomonas aeruginosa.  
XX  
PN WO200170955-A2.  
XX  
XX 27-SEP-2001.  
XX  
XX 21-MAR-2001; 2001WO-US09180.  
XX  
XX 21-MAR-2000; 2000US-191078P.  
XX 23-MAY-2000; 2000US-206848P.  
XX 26-MAY-2000; 2000US-20727P.  
XX 23-OCT-2000; 2000US-242578P.  
XX 27-NOV-2000; 2000US-253625P.  
XX 22-DEC-2000; 2000US-257931P.  
XX 16-FEB-2001; 2001US-269308P.  
XX  
XX (ELIT-) ELITRA PHARM INC.  
XX  
XX Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GU;  
XX Yamamoto RT, Xu HH;  
XX  
XX WPI: 2001-611495/70.  
XX P-PSDB; AAU36417.  
XX  
XX New polynucleotides for the identification and development of  
XX antibiotics, comprise sequences of antisense nucleic acids -  
XX  
XX Claim 27; Seq ID No 7913; 511pp; English.  
XX  
XX The invention relates to antisense inhibitors of genes essential to  
CC prokaryotic cellular proliferation, their use in identifying the  
CC genes, their use in the discovery of novel antibiotics, the essential  
CC genes themselves and the encoded proteins. The prokaryotes used are  
CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella  
CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The  
CC invention is also useful for the identification of potential new targets  
CC for antibiotic development. The antisense nucleic acids can also be used  
CC to identify proteins used in proliferation, to express these proteins,  
CC and to obtain antibodies capable of binding to the expressed proteins.  
CC The proteins can be used to screen compounds in rational drug discovery  
CC programmes. The antisense nucleic acid sequence is also useful to screen  
CC for homologous nucleic acids which are required for cell proliferation in  
CC a wide variety of organisms. The present sequence encodes an  
CC essential prokaryotic cellular proliferation protein.  
CC Note: The sequence data for this patent did not form part  
CC of the printed specification, but was obtained in electronic  
CC format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
XX Sequence 1083 BP; 163 A; 348 C; 336 G; 236 T; 0 other;  
SQ  
Query Match 2.5%; Score 36; DB 23; Length 1083;  
Best Local Similarity 100.0%; Pred. No. 2.6e-06;  
Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 2 GTGCTGATGGCGCTCGCCACCTTGAAGTCGCTTGA 37  
|||||  
Db 1048 GTGCTGATGGCGCTCGCCACCTTGAAGTCGCTTGA 1083  
RESULT 7  
AAS89258  
ID AAS89258 standard; cDNA; 1080 BP.  
XX  
AC AAS89258;  
XX  
XX 13-FEB-2002 (first entry)  
XX  
XX

DE DNA encoding novel human diagnostic protein #25062.  
 XX Human; chromosome mapping; gene mapping; gene therapy; forensic;  
 KW food supplement; medical imaging; diagnostic; genetic disorder; ss.  
 KW Homo sapiens.  
 OS  
 XX WO200175067-A2.  
 PN 11-OCT-2001.  
 PD 30-MAR-2001; 2001WO-US08631.  
 XX 31-MAR-2000; 2000US-0540217.  
 PF 23-AUG-2000; 2000US-0649167.  
 XX (HYSE-) HYSEQ INC.  
 PA Drmanac RT, Liu C, Tang YT;  
 XX WPI; 2001-639362/73.  
 PI P-PSDB; ABG25071.  
 DR  
 DR New isolated polynucleotide and encoded polypeptides, useful in  
 XX diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity  
 PT  
 XX Claim 1; SEQ ID No 25062; 103pp; English.  
 PS  
 XX The invention relates to isolated polynucleotide (I) and  
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
 CC and gene mapping, and in recombinant production of (II). The  
 CC polynucleotides are also used in diagnostics as expressed sequence tags  
 CC for identifying expressed genes. (I) is useful in gene therapy techniques  
 CC to restore normal activity of (II) or to treat disease states involving  
 CC quantitating a polypeptide in tissue, as molecular weight markers and as  
 CC a food supplement. (II) and its binding partners are useful in medical  
 CC imaging of sites expressing (II). (I) and (II) are useful for treating  
 CC disorders involving aberrant protein expression or biological activity.  
 CC The polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. AAS64197-AAS94564 represent novel human  
 CC diagnostic coding sequences of the invention.  
 CC Note: The sequence data for this patent did not appear in the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 XX Sequence 1080 BP; 267 A; 253 C; 299 G; 261 T; 0 other;  
 SQ  
 Query Match 1.7%; Score 24; DB 23; Length 1080;  
 Best Local Similarity 100.0%; Pred. No. 0.67;  
 Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 537 CTGTCGAGCTTCAGCTGGAAACC 560  
 |||||  
 Db 40 CTGTCGAGCTTCAGCTGGAAACC 63  
 RESULT 8  
 AAS89801  
 ID AAS89801 standard; cDNA; 1080 BP.  
 XX  
 AC AAS89801;  
 XX 13-FEB-2002 (first entry)  
 DT  
 DE DNA encoding novel human diagnostic protein #25062.  
 XX Human; chromosome mapping; gene mapping; gene therapy; forensic;  
 KW food supplement; medical imaging; diagnostic; genetic disorder; ss.  
 KW Homo sapiens.  
 OS  
 XX WO200175067-A2.  
 PN 11-OCT-2001.  
 PD 30-MAR-2001; 2001WO-US08631.  
 XX 31-MAR-2000; 2000US-0540217.  
 PF 23-AUG-2000; 2000US-0649167.  
 XX (HYSE-) HYSEQ INC.  
 PA Drmanac RT, Liu C, Tang YT;  
 XX WPI; 2001-639362/73.  
 PI P-PSDB; ABG25071.  
 DR  
 DR New isolated polynucleotide and encoded polypeptides, useful in  
 XX diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity  
 PT  
 XX Claim 1; SEQ ID No 25062; 103pp; English.  
 PS  
 XX The invention relates to isolated polynucleotide (I) and  
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
 CC and gene mapping, and in recombinant production of (II). The  
 CC polynucleotides are also used in diagnostics as expressed sequence tags  
 CC for identifying expressed genes. (I) is useful in gene therapy techniques  
 CC to restore normal activity of (II) or to treat disease states involving  
 CC quantitating a polypeptide in tissue, as molecular weight markers and as  
 CC a food supplement. (II) and its binding partners are useful in medical  
 CC imaging of sites expressing (II). (I) and (II) are useful for treating  
 CC disorders involving aberrant protein expression or biological activity.  
 CC The polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. AAS64197-AAS94564 represent novel human  
 CC diagnostic coding sequences of the invention.  
 CC Note: The sequence data for this patent did not appear in the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 XX Sequence 1080 BP; 267 A; 253 C; 299 G; 261 T; 0 other;  
 SQ  
 Query Match 1.7%; Score 24; DB 23; Length 1080;  
 Best Local Similarity 100.0%; Pred. No. 0.67;  
 Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 537 CTGTCGAGCTTCAGCTGGAAACC 560  
 |||||  
 Db 40 CTGTCGAGCTTCAGCTGGAAACC 63  
 RESULT 8  
 AAS89801  
 ID AAS89801 standard; cDNA; 1080 BP.  
 XX  
 AC AAS89801;  
 XX 13-FEB-2002 (first entry)  
 DT  
 DE DNA encoding novel human diagnostic protein #25062.  
 XX Human; chromosome mapping; gene mapping; gene therapy; forensic;  
 KW food supplement; medical imaging; diagnostic; genetic disorder; ss.  
 KW Homo sapiens.  
 OS  
 XX WO200175067-A2.  
 PN 11-OCT-2001.  
 PD 30-MAR-2001; 2001WO-US08631.  
 XX 31-MAR-2000; 2000US-0540217.  
 PF 23-AUG-2000; 2000US-0649167.  
 XX (HYSE-) HYSEQ INC.  
 PA Drmanac RT, Liu C, Tang YT;  
 XX WPI; 2001-639362/73.  
 PI P-PSDB; ABG25071.  
 DR  
 DR New isolated polynucleotide and encoded polypeptides, useful in  
 XX diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity  
 PT  
 XX Claim 1; SEQ ID No 25062; 103pp; English.  
 PS  
 XX The invention relates to isolated polynucleotide (I) and  
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
 CC and gene mapping, and in recombinant production of (II). The  
 CC polynucleotides are also used in diagnostics as expressed sequence tags  
 CC for identifying expressed genes. (I) is useful in gene therapy techniques  
 CC to restore normal activity of (II) or to treat disease states involving  
 CC quantitating a polypeptide in tissue, as molecular weight markers and as  
 CC a food supplement. (II) and its binding partners are useful in medical  
 CC imaging of sites expressing (II). (I) and (II) are useful for treating  
 CC disorders involving aberrant protein expression or biological activity.  
 CC The polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. AAS64197-AAS94564 represent novel human  
 CC diagnostic coding sequences of the invention.  
 CC Note: The sequence data for this patent did not appear in the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 XX Sequence 1080 BP; 267 A; 253 C; 299 G; 261 T; 0 other;  
 SQ  
 Query Match 1.7%; Score 24; DB 23; Length 1080;  
 Best Local Similarity 100.0%; Pred. No. 0.67;  
 Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 537 CTGTCGAGCTTCAGCTGGAAACC 560  
 |||||  
 Db 40 CTGTCGAGCTTCAGCTGGAAACC 63  
 RESULT 9  
 AAS90166  
 ID AAS90166 standard; cDNA; 1080 BP.  
 XX  
 AC AAS90166;  
 XX 13-FEB-2002 (first entry)  
 DT  
 DE DNA encoding novel human diagnostic protein #25970.  
 XX Human; chromosome mapping; gene mapping; gene therapy; forensic;  
 KW food supplement; medical imaging; diagnostic; genetic disorder; ss.

KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
 KW food supplement; medical imaging; diagnostic; genetic disorder; ss.  
 OS Homo sapiens.  
 XX WO200175067-A2.  
 PN 11-OCT-2001.  
 PD 30-MAR-2001; 2001WO-US08631.  
 XX 31-MAR-2000; 2000US-0540217.  
 PF 23-AUG-2000; 2000US-0649167.  
 XX (HYSE-) HYSEQ INC.  
 PA Drmanac RT, Liu C, Tang YT;  
 XX WPI; 2001-639362/73.  
 PI P-PSDB; ABG25614.  
 DR  
 DR New isolated polynucleotide and encoded polypeptides, useful in  
 XX diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity  
 PT  
 XX Claim 1; SEQ ID No 25605; 103pp; English.  
 PS  
 XX The invention relates to isolated polynucleotide (I) and  
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
 CC and gene mapping, and in recombinant production of (II). The  
 CC polynucleotides are also used in diagnostics as expressed sequence tags  
 CC for identifying expressed genes. (I) is useful in gene therapy techniques  
 CC to restore normal activity of (II) or to treat disease states involving  
 CC quantitating a polypeptide in tissue, as molecular weight markers and as  
 CC a food supplement. (II) and its binding partners are useful in medical  
 CC imaging of sites expressing (II). (I) and (II) are useful for treating  
 CC disorders involving aberrant protein expression or biological activity.  
 CC The polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. AAS64197-AAS94564 represent novel human  
 CC diagnostic coding sequences of the invention.  
 CC Note: The sequence data for this patent did not appear in the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 XX Sequence 1080 BP; 267 A; 253 C; 299 G; 261 T; 0 other;  
 SQ  
 Query Match 1.7%; Score 24; DB 23; Length 1080;  
 Best Local Similarity 100.0%; Pred. No. 0.67;  
 Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 537 CTGTCGAGCTTCAGCTGGAAACC 560  
 |||||  
 Db 40 CTGTCGAGCTTCAGCTGGAAACC 63  
 RESULT 9  
 AAS90166  
 ID AAS90166 standard; cDNA; 1080 BP.  
 XX  
 AC AAS90166;  
 XX 13-FEB-2002 (first entry)  
 DT  
 DE DNA encoding novel human diagnostic protein #25970.  
 XX Human; chromosome mapping; gene mapping; gene therapy; forensic;  
 KW food supplement; medical imaging; diagnostic; genetic disorder; ss.

```

XX OS Homo sapiens.
XX PN WO200175067-A2.
XX PD 11-OCT-2001.
XX PF 30-MAR-2001; 2001WO-US08631.
XX PR 31-MAR-2000; 2000US-0540217.
XX PR 23-AUG-2000; 2000US-0649167.
XX PA (HYSE-) HYSEQ INC.
XX PI Drmanac RT, Liu C, Tang YT;
XX DR WPI; 2001-639362/73.
XX DR P-PSDB; ABG25979.
XX PT New isolated polynucleotide and encoded polypeptides, useful in
XX PT diagnostics, forensics, gene mapping, identification of mutations
XX PT responsible for genetic disorders or other traits and to assess
XX PT biodiversity -
XX PS Claim 1; SEQ ID No 25970; 103pp; English.
XX CC The invention relates to isolated polynucleotide (I) and
XX CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
XX CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
XX CC and gene mapping, and in recombinant production of (II). The
XX CC polynucleotides are also used in diagnostics as expressed sequence tags
XX CC for identifying expressed genes. (I) is useful in gene therapy techniques
XX CC to restore normal activity of (II) or to treat disease states involving
XX CC quantitating a polypeptide in tissue, as molecular weight markers and as
XX CC a food supplement. (II) and its binding partners are useful in medical
XX CC imaging of sites expressing (II). (I) and (II) are useful for treating
XX CC disorders involving aberrant protein expression or biological activity.
XX CC The polypeptide and polynucleotide sequences have applications in
XX CC diagnostics, forensics, gene mapping, identification of mutations
XX CC responsible for genetic disorders or other traits to assess biodiversity
XX CC and to produce other types of data and products dependent on DNA and
XX CC amino acid sequences. AAS64197-AAS94564 represent novel human
XX CC diagnostic coding sequences of the invention.
XX CC Note: The sequence data for this patent did not appear in the printed
XX CC specification, but was obtained in electronic format directly from WIPO
XX CC at ftp.wipo.int/pub/published_pct_sequences.
XX SQ Sequence 1080 BP; 267 A; 253 C; 299 G; 261 T; 0 other;

Query Match 1.7%; Score 24; DB 23; Length 1080;
Best Local Similarity 100.0%; Pred. No. 0.67;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 537 CTGTCGAGCTTCAGCTGGAAC 560
DB 40 CTGTCGAGCTTCAGCTGGAAC 63
|||||
RESULT 10
AAH81460
ID AAH81460 standard; DNA; 1317 BP.
XX AC AAH81460;
XX DT 21-SEP-2001 (first entry)
XX DE Escherichia coli protein encoding nucleotide sequence SEQ ID NO:259.
XX KW Escherichia coli; identification; proliferation; microorganism;
XX KW antimicrobial; antibacterial; antibiotic; gene therapy; diagnosis;
XX KW bacterial growth inhibition; ds.
XX

```

```

OS Escherichia coli.
XX PN WO200148209-A2.
XX PD 05-JUL-2001.
XX PF 19-DEC-2000; 2000WO-US34419.
XX PR 23-DEC-1999; 99US-0173005.
XX PA (ELIT-) ELITRA PHARM INC.
XX PI Forsyth RA, Ohlsen KL, Zyskind JW;
XX DR WPI; 2001-457376/49.
XX DR P-PSDB; AAG98404.
XX PT Novel nucleic acids encoding proteins required for Escherichia coli
XX PT proliferation, useful for screening for antimicrobial agents -
XX PS Claim 9; Page 388-389; 596pp; English.
XX CC The present invention describes a purified or isolated nucleic acid
XX CC sequence (I) consisting essentially of one of the 93 nucleotide sequences
XX CC given in AAH81202 to AAH81294, where expression of the nucleic acid in a
XX CC microorganism is capable of inhibiting proliferation of a microorganism.
XX CC (I) have antibacterial and antibiotic activities, and can be used in
XX CC gene therapy. Expression of (I) in a microorganism inhibits proliferation
XX CC of the microorganism, and the manufactured antibiotic is useful for
XX CC reducing the activity or level of a gene product required for
XX CC proliferation of a microorganism in a subject, specifically humans. The
XX CC nucleic acids that inhibit bacterial growth or proliferation can be used
XX CC as antisense therapeutics for killing bacteria. In addition to
XX CC therapeutic applications, the nucleic acid sequences complementary to
XX CC sequences required for proliferation can be used as diagnostic tools.
XX CC For example, nucleic acid probes complementary to proliferation-required
XX CC sequences that are specific for particular species of microorganisms can
XX CC be used as probes to identify particular microorganism species in
XX CC clinical specimens. AAH81295 to AAH81487 encode the Escherichia coli
XX CC proteins given in AAG98239 to AAG98431, and AAH81488 to AAH81491
XX CC represent oligonucleotides, which are used in the exemplification of the
XX CC present invention.
XX SQ Sequence 1317 BP; 294 A; 323 C; 400 G; 300 T; 0 other;

Query Match 1.7%; Score 24; DB 22; Length 1317;
Best Local Similarity 100.0%; Pred. No. 0.66;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 537 CTGTCGAGCTTCAGCTGGAAC 560
DB 475 CTGTCGAGCTTCAGCTGGAAC 498
|||||
RESULT 11
AAS52295
ID AAS52295 standard; DNA; 1317 BP.
XX AC AAS52295;
XX DT 13-FEB-2002 (first entry)
XX DE E. coli DNA for cellular proliferation protein #17.
XX KW Antisense; ds; prokaryotic cellular proliferation gene;
XX KW antibiotic; antibacterial; drug design.
XX OS Escherichia coli.
XX PN WO200170955-A2.
XX PD 27-SEP-2001.
XX

```

```

PF 21-MAR-2001; 2001WO-US09180.
XX
XX 21-MAR-2000; 2000US-191078P.
XX 23-MAY-2000; 2000US-206848P.
XX 26-MAY-2000; 2000US-207727P.
XX 23-OCT-2000; 2000US-242578P.
XX 27-NOV-2000; 2000US-253625P.
XX 22-DEC-2000; 2000US-257931P.
XX 16-FEB-2001; 2001US-269308P.
XX
XX (ELIT-) ELITRA PHARM INC.
XX
XX Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;
XX Yamamoto RT, Xu HH;
XX WPI: 2001-611495/70.
XX P-PSDB; AAU34436.
XX
XX New polynucleotides for the identification and development of
XX antibiotics, comprise sequences of antisense nucleic acids -
XX
XX Claim 27; Seq ID No 5932; 511pp; English.
XX
XX The invention relates to antisense inhibitors of genes essential to
XX prokaryotic cellular proliferation, their use in identifying the
XX genes, their use in the discovery of novel antibiotics, the essential
XX genes themselves and the encoded proteins. The prokaryotes used are
XX Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
XX pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
XX invention is also useful for the identification of potential new targets
XX for antibiotic development. The antisense nucleic acids can also be used
XX to identify proteins used in proliferation, to express these proteins,
XX and to obtain antibodies capable of binding to the expressed proteins.
XX The proteins can be used to screen compounds in rational drug discovery
XX programmes. The antisense nucleic acid sequence is also useful to screen
XX for homologous nucleic acids which are required for cell proliferation in
XX a wide variety of organisms. The present sequence encodes an
XX essential prokaryotic cellular proliferation protein.
XX Note: The sequence data for this patent did not form part
XX of the printed specification, but was obtained in electronic
XX format directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 1317 BP; 294 A; 323 C; 400 G; 300 T; 0 other;
XX
XX Query Match 1.7%; Score 24; DB 23; Length 1317;
XX Best Local Similarity 100.0%; Pred. No. 0.66;
XX Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 537 CTGTCGAGCTTCACGCTGGAAC 560
XX |||||||||||||||||||
XX DB 475 CTGTCGAGCTTCACGCTGGAAC 498
XX
XX RESULT 12
XX AAS73012
XX ID AAS73012 standard; cDNA; 2058 BP.
XX
XX AC AAS73012;
XX
XX DT 13-FEB-2002 (first entry)
XX
XX DE DNA encoding novel human diagnostic protein #8816.
XX
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
XX OS Homo sapiens.
XX
XX PN WO200175067-A2.
XX
XX PD 11-OCT-2001.
XX
XX PF 30-MAR-2001; 2001WO-US08631.
XX
XX
XX 30-MAR-2001; 2001WO-US08631.
XX
XX 31-MAR-2000; 2000US-0540217.
XX 23-AUG-2000; 2000US-0649167.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Drmanac RT, Liu C, Tang YT;
XX
XX WPI: 2001-639362/73.
XX P-PSDB; ABG08825.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits and to assess
XX biodiversity -
XX
XX Claim 1; SEQ ID No 8816; 103pp; English.
XX
XX The invention relates to isolated polynucleotide (I) and
XX polypeptide (II) sequences. (I) is useful as hybridisation probes,
XX polymerase chain reaction (PCR) primers, oligomers, and for chromosome
XX and gene mapping, and in recombinant production of (II). The
XX polynucleotides are also used in diagnostics as expressed sequence tags
XX for identifying expressed genes. (I) is useful in gene therapy techniques
XX to restore normal activity of (II) or to treat disease states involving
XX (II). (II) is useful for generating antibodies against it, detecting or
XX quantitating a polypeptide in tissue, as molecular weight markers and as
XX a food supplement. (II) and its binding partners are useful in medical
XX imaging of sites expressing (II). (I) and (II) are useful for treating
XX disorders involving aberrant protein expression or biological activity.
XX The polypeptide and polynucleotide sequences have applications in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits to assess biodiversity
XX and to produce other types of data and products dependent on DNA and
XX amino acid sequences. AAS64197-AAS94564 represent novel human
XX diagnostic coding sequences of the invention.
XX Note: The sequence data for this patent did not appear in the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 2058 BP; 496 A; 526 C; 574 G; 462 T; 0 other;
XX
XX Query Match 1.7%; Score 24; DB 23; Length 2058;
XX Best Local Similarity 100.0%; Pred. No. 0.63;
XX Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 537 CTGTCGAGCTTCACGCTGGAAC 560
XX |||||||||||||||||||
XX DB 1801 CTGTCGAGCTTCACGCTGGAAC 1824
XX
XX RESULT 13
XX AAS69095
XX ID AAS69095 standard; cDNA; 3117 BP.
XX
XX AC AAS69095;
XX
XX DT 13-FEB-2002 (first entry)
XX
XX DE DNA encoding novel human diagnostic protein #4899.
XX
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
XX OS Homo sapiens.
XX
XX PN WO200175067-A2.
XX
XX PD 11-OCT-2001.
XX
XX PF 30-MAR-2001; 2001WO-US08631.
XX
XX

```



PR 31-MAR-2000; 2000US-0540217.  
 XX 23-AUG-2000; 2000US-0649167.  
 PA (HYSE-) HYSEQ INC.  
 XX Drmanac RT, Liu C, Tang YT;  
 XX WPI; 2001-639362/73.  
 DR P-PSDB; ABG04908.  
 XX New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity -  
 XX Claim 1; SEQ ID No 4899; 103pp; English.  
 XX The invention relates to isolated polynucleotide (I) and  
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
 CC and gene mapping, and in recombinant production of (II). The  
 CC polynucleotides are also used in diagnostics as expressed sequence tags  
 CC for identifying expressed genes. (I) is useful in gene therapy techniques  
 CC to restore normal activity of (II) or to treat disease states involving  
 CC quantitating a polypeptide in tissue, as molecular weight markers and as  
 CC a food supplement. (II) and its binding partners are useful in medical  
 CC imaging of sites expressing (II). (I) and (II) are useful for treating  
 CC disorders involving aberrant protein expression or biological activity.  
 CC The polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. AAS64197-AAS94564 represent novel human  
 CC diagnostic coding sequences of the invention.  
 CC Note: The sequence data for this patent did not appear in the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX SQ Sequence 3117 BP; 749 A; 820 C; 856 G; 692 T; 0 other;  
 Query Match 1.7%; Score 24; DB 23; Length 3117;  
 Best Local Similarity 100.0%; Pred. No. 0.6;  
 Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 537 CTGTCGAGCTTCACGCTGGAACC 560  
 Db |||||||||||||||||||  
 373 CTGTCGAGCTTCACGCTGGAACC 396  
 RESULT 14  
 AAS89240  
 ID AAS89240 standard; cDNA; 3117 BP.  
 XX AC AAS89240;  
 XX DT 13-FEB-2002 (first entry)  
 XX DNA encoding novel human diagnostic protein #25044.  
 DE Human; chromosome mapping; gene mapping; gene therapy; forensic;  
 DE food supplement; medical imaging; diagnostic; genetic disorder; ss.  
 KW Homo sapiens.  
 XX OS WO200175067-A2.  
 XX PN 11-OCT-2001.  
 XX PD 30-MAR-2001; 2001WO-US08631.  
 XX PF 31-MAR-2000; 2000US-0540217.  
 XX PR 31-MAR-2000; 2000US-0540217.  
 XX PR 23-AUG-2000; 2000US-0649167.  
 XX PA (HYSE-) HYSEQ INC.

XX (HYSE-) HYSEQ INC.  
 XX Drmanac RT, Liu C, Tang YT;  
 XX WPI; 2001-639362/73.  
 DR P-PSDB; ABG25053.  
 XX New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity -  
 XX Claim 1; SEQ ID No 25044; 103pp; English.  
 XX The invention relates to isolated polynucleotide (I) and  
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
 CC and gene mapping, and in recombinant production of (II). The  
 CC polynucleotides are also used in diagnostics as expressed sequence tags  
 CC for identifying expressed genes. (I) is useful in gene therapy techniques  
 CC to restore normal activity of (II) or to treat disease states involving  
 CC quantitating a polypeptide in tissue, as molecular weight markers and as  
 CC a food supplement. (II) and its binding partners are useful in medical  
 CC imaging of sites expressing (II). (I) and (II) are useful for treating  
 CC disorders involving aberrant protein expression or biological activity.  
 CC The polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. AAS64197-AAS94564 represent novel human  
 CC diagnostic coding sequences of the invention.  
 CC Note: The sequence data for this patent did not appear in the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX SQ Sequence 3117 BP; 749 A; 820 C; 856 G; 692 T; 0 other;  
 Query Match 1.7%; Score 24; DB 23; Length 3117;  
 Best Local Similarity 100.0%; Pred. No. 0.6;  
 Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 537 CTGTCGAGCTTCACGCTGGAACC 560  
 Db |||||||||||||||||||  
 373 CTGTCGAGCTTCACGCTGGAACC 396  
 RESULT 15  
 AAS94542  
 ID AAS94542 standard; cDNA; 4831 BP.  
 XX AC AAS94542;  
 XX DT 13-FEB-2002 (first entry)  
 XX DNA encoding novel human diagnostic protein #30346.  
 DE Human; chromosome mapping; gene mapping; gene therapy; forensic;  
 DE food supplement; medical imaging; diagnostic; genetic disorder; ss.  
 KW Homo sapiens.  
 XX OS WO200175067-A2.  
 XX PN 11-OCT-2001.  
 XX PD 30-MAR-2001; 2001WO-US08631.  
 XX PF 31-MAR-2000; 2000US-0540217.  
 XX PR 23-AUG-2000; 2000US-0649167.  
 XX PR (HYSE-) HYSEQ INC.

```

XX Drmanac RT, Liu C, Tang YT;
XX WPI: 2001-639362/73.
XX P-PSDB: ABG30355.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits and to assess
XX biodiversity
XX
XX Claim 1: SEQ ID No 30346; 103pp; English.
XX
XX The invention relates to isolated polynucleotide (I) and
XX polypeptide (II) sequences. (I) is useful as hybridisation probes,
XX polymerase chain reaction (PCR) primers, oligomers, and for chromosome
XX and gene mapping, and in recombinant production of (II). The
XX polynucleotides are also used in diagnostics as expressed sequence tags
XX for identifying expressed genes. (I) is useful in gene therapy techniques
XX to restore normal activity of (II) or to treat disease states involving
XX (II). (II) is useful for generating antibodies against it, detecting or
XX quantitating a polypeptide in tissue, as molecular weight markers and as
XX a food supplement. (II) and its binding partners are useful in medical
XX imaging of sites expressing (II). (I) and (II) are useful for treating
XX disorders involving aberrant protein expression or biological activity.
XX The polypeptide and polynucleotide sequences have applications in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits to assess biodiversity
XX and to produce other types of data and products dependent on DNA and
XX amino acid sequences. AAS64197-RAS94564 represent novel human
XX diagnostic coding sequences of the invention.
XX Note: The sequence data for this patent did not appear in the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 4831 BP; 1180 A; 1249 C; 1353 G; 1049 T; 0 other;
XX
XX Query Match 1.7%; Score 24; DB 23; Length 4831;
XX Best Local Similarity 100.0%; Pred. No. 0.58;
XX Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 537 CTGTCGAGCTTCAGCTGGAACC 560
XX Db 4574 CTGTCGAGCTTCAGCTGGAACC 4597
XX
XX Search completed: August 14, 2003, 04:47:00
XX Job time : 425 secs

```

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 14, 2003, 01:48:19 ; Search time 1500 Seconds  
(without alignments)  
2163.342 Million cell updates/sec

Title: us-09-701-229-1

Perfect score: 1450

Sequence: 1 cgtgctgctgcctgcgcga.....tcttgagcggcgacggcgc 1450

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1504479 seqs, 1118970152 residues

Total number of hits satisfying chosen parameters: 3008958

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published\_Applications\_NA.:

1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq.\*  
2: /cgn2\_6/ptodata/1/pubpna/PCT\_NEW\_PUB.seq.\*  
3: /cgn2\_6/ptodata/1/pubpna/US06\_NEW\_PUB.seq.\*  
4: /cgn2\_6/ptodata/1/pubpna/US06\_PUBCOMB.seq.\*  
5: /cgn2\_6/ptodata/1/pubpna/US07\_NEW\_PUB.seq.\*  
6: /cgn2\_6/ptodata/1/pubpna/PCTUS\_PUBCOMB.seq.\*  
7: /cgn2\_6/ptodata/1/pubpna/US08\_NEW\_PUB.seq.\*  
8: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq.\*  
9: /cgn2\_6/ptodata/1/pubpna/US09A\_PUBCOMB.seq.\*  
10: /cgn2\_6/ptodata/1/pubpna/US09B\_PUBCOMB.seq.\*  
11: /cgn2\_6/ptodata/1/pubpna/US09C\_PUBCOMB.seq.\*  
12: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq.\*  
13: /cgn2\_6/ptodata/1/pubpna/US10A\_PUBCOMB.seq.\*  
14: /cgn2\_6/ptodata/1/pubpna/US10B\_PUBCOMB.seq.\*  
15: /cgn2\_6/ptodata/1/pubpna/US10\_NEW\_PUB.seq.\*  
16: /cgn2\_6/ptodata/1/pubpna/US50\_NEW\_PUB.seq.\*  
17: /cgn2\_6/ptodata/1/pubpna/US60\_PUBCOMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Query	Score	Length	ID	Description
1	1347	92.9	1347	9	US-09-815-242-7912
2	332.8	23.0	1317	9	US-09-815-242-9978
3	278.6	19.2	1317	9	US-09-741-669-259
4	278.6	19.2	1317	9	US-09-815-242-5932
5	180.6	12.5	9025608	14	US-10-156-761-1
6	180.2	12.4	1401	14	US-10-156-761-6097
7	173.8	12.0	1461	9	US-09-712-363-79
8	149.8	10.3	1314	9	US-09-815-242-7090
9	149.8	10.3	1830121	14	US-10-329-360-1
10	91.8	6.3	1352	10	US-09-974-300-1685
c 11	86	5.9	3309400	10	US-09-738-626-1
12	68.4	4.7	795	14	US-10-156-761-5987
c 13	68.4	4.7	9025608	14	US-10-156-761-1
c 14	67	4.6	640681	10	US-09-790-988-1
15	66.2	4.6	3759	14	US-10-156-761-2268
16	66	4.6	1959	14	US-10-156-761-2777

17	66	4.6	88421	10	US-09-815-242-7912	Sequence 1, Appli
18	65.8	4.5	786	14	US-10-187-267A-32	Sequence 32, Appli
19	65.8	4.5	36321	14	US-10-187-267A-1	Sequence 1, Appli
20	65.6	4.5	1134	14	US-10-156-761-3946	Sequence 3946, Ap
21	64.4	4.4	1458	14	US-10-156-761-4634	Sequence 4634, Ap
22	64.4	4.4	1620	14	US-10-156-761-6900	Sequence 6900, Ap
23	62.8	4.3	1392	14	US-10-156-761-2193	Sequence 2193, Ap
24	62.4	4.3	2214	14	US-10-156-761-2510	Sequence 2510, Ap
25	62.2	4.3	1533	9	US-09-815-242-7755	Sequence 7755, Ap
c 26	62	4.3	5970	11	US-09-793-708-21	Sequence 21, Appli
c 27	62	4.3	5970	12	US-10-201-365-11	Sequence 11, Appli
28	61.4	4.2	1197	9	US-09-815-242-7683	Sequence 7683, Ap
29	61.4	4.2	1848	14	US-10-156-761-1880	Sequence 1880, Ap
30	60.8	4.2	1248	10	US-09-861-289-7	Sequence 7, Appli
31	60.8	4.2	1248	10	US-09-860-846-7	Sequence 7, Appli
32	60.8	4.2	1248	11	US-09-988-384B-7	Sequence 7, Appli
33	60.8	4.2	1248	11	US-09-836-821-7	Sequence 7, Appli
c 34	60.8	4.2	12441	11	US-09-988-384B-3	Sequence 3, Appli
c 35	60.8	4.2	13613	10	US-09-861-289-3	Sequence 3, Appli
c 36	60.8	4.2	13613	10	US-09-860-846-3	Sequence 3, Appli
c 37	60.8	4.2	13613	11	US-09-836-821-3	Sequence 3, Appli
38	59.6	4.1	15872	10	US-09-861-289-1	Sequence 1, Appli
39	59.6	4.1	15872	10	US-09-860-846-1	Sequence 1, Appli
40	59.6	4.1	15872	11	US-09-988-384B-1	Sequence 1, Appli
41	59.6	4.1	15872	11	US-09-836-821-1	Sequence 1, Appli
42	59.2	4.1	1167	14	US-10-156-761-6049	Sequence 6049, Ap
43	59	4.1	966	14	US-10-156-761-5269	Sequence 5269, Ap
c 44	58.8	4.1	1092	14	US-10-156-761-1710	Sequence 1710, Ap
45	58.8	4.1	1545	12	US-09-953-348-49	Sequence 49, Appli

#### ALIGNMENTS

RESULT 1  
US-09-815-242-7912  
Sequence 7912, Application US/09815242  
Patent No. US20020061569A1

GENERAL INFORMATION:  
APPLICANT: Haseibeck, Robert  
APPLICANT: Ohlsen, Karl L.  
APPLICANT: Zyskind, Judith W.  
APPLICANT: Wall, Daniel  
APPLICANT: Trawick, John D.  
APPLICANT: Carr, Grant J.  
APPLICANT: Yamamoto, Robert T.  
APPLICANT: Xu, H. Howard

TITLE OF INVENTION: Identification of Essential Genes in  
FILE OF INVENTION: Prokaryotes

FILE REFERENCE: ELITRA.011A

CURRENT APPLICATION NUMBER: US/09/815,242

PRIOR FILING DATE: 2001-03-21

PRIOR APPLICATION NUMBER: 60/191,078

PRIOR FILING DATE: 2000-03-21

PRIOR APPLICATION NUMBER: 60/206,848

PRIOR FILING DATE: 2000-05-23

PRIOR APPLICATION NUMBER: 60/207,727

PRIOR FILING DATE: 2000-05-26

PRIOR APPLICATION NUMBER: 60/242,578

PRIOR FILING DATE: 2000-10-23

PRIOR APPLICATION NUMBER: 60/253,625

PRIOR FILING DATE: 2000-11-27

PRIOR APPLICATION NUMBER: 60/257,931

PRIOR FILING DATE: 2000-12-22

PRIOR APPLICATION NUMBER: 60/269,308

PRIOR FILING DATE: 2001-02-16

NUMBER OF SEQ ID NOS: 14110

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 7912

LENGTH: 1347

TYPE: DNA

ORGANISM: Pseudomonas aeruginosa

FEATURE:



```
Matches 747; Conservative 0; Mismatches 552; Indels 33; Gaps 5;
QY 69 GACCACTTCGGCATCTGTCGCGCTCGCAAGAGCGGATGCCCTGGTGGCGCTACCTG 128
Db 16 GACAAAACGTCGTCATATCGGCTCGGGCTTAACCGGACTCTCTTGGGTGGACTTTTT 75
QY 129 GCGCGCGGGGCTTGGCTTTTCGCGGTGGTCGATACCCGAGAGAACCGCGGAGCTGGCC 188
Db 76 CTGCGCCGCGGCTGACGCGCGGGGTGATGGATACCTCGGTGACGCGCGGGTCTGGAT 135
QY 189 ACCCTCGTCCGATATCGGAGGTGGAGTGGCGGTGGCGGAACTCGACGCCGAGTTC 248
Db 136 AAGCTCCG-----CAAGAGTTGAGCGTACGTTGGCGGCTGAACGACGAGTGG 186
QY 249 CTCTGCTCGCGCGGCAACTCTATGTCAGCCCGGCTTGTGCTGCGCGCACCCCTCGCTG 308
Db 187 CTCCTAGCGGGGATTAATCGCTCGCGAGCCCTGGTATGCTCTGGCGATCCCTCTCTT 246
QY 309 GTACAGCGCGCGCGAAAGCGGTGCGCATCTCCGGTGACATCGATCTTTCGCCCGCGAG 368
Db 247 AGCGCTGCGCGCCAGCGCTGGAGTGGAAATTGTGCGGATATCGAACTGTTTTCGCCGGA 306
QY 369 GCGAAGCGCGCGATCTCGCCATCACCGGTTTCCAAACGGAAGACACCGTACCACCTG 428
Db 307 GCGAAGCGCGGATTTGTGCCATCACCGGCTCGAACGGAAGAACCGCTGACACCTTA 366
QY 429 GTGGCGAAATGGCGGTGGCGGCAAGCGGTGCGCGTGGCGGCAACCTCGCGCAC 488
Db 367 GTGGCGAGATGCGAAGCGGGCGGCTCAATGTGCGGTGGCGGCAATATCGTCTT 426
QY 489 CGGGCGCTCGACTGTGCGCGGACGATCGAGCTGTAGCTGTGGAGTGTGCGAGTTC 548
Db 427 CGGGCGCTGATGCTGTGGATCGGATCGCAATTTAGCTGTGGAATATCCAGTTTC 486
QY 549 CAGCTGGAACCTGCGATCGCTCAAGCGGAGGTGGCGACCGTGTGAACGTCAGCGAA 608
Db 487 CAATTGGAACCTACCTCAAGTTTGAAGCGGCGGCAACGTTGTCAACGTCACCTGAA 546
QY 609 GACCATATGATCGCTA---CGAGCGCATGGCTGACTACACCTGGCCAGACACCGATC 665
Db 547 GATCATATGAGCGGTATCGTTTGTGTTTCCACACTACCGGCGCGGAACCTGGCGTC 606
QY 666 TTCGGGTTGCGCGCGAGTGTGGTGAATGCGCGCGATGCCCTGACCCGACCGCTGATC 725
Db 607 TAGGAGAAGCGAAGTGTGCTGAGTGAATGCGGATGCGGATGCGGATGCGGATGCG 666
QY 726 GCGGATACCGTGGCTGGTGGCTGCAACAGCGGACTTCAGGCTTTCGGC 785
Db 667 GGGCGGATGAGCGCTGCGTACGCTTGGCGTCAATATGGGTGATTATCA-----C 717
QY 786 CTGATCGAGAAGACGCCAGAAAGTGGCTGGCGTTCCAGTTCCAGAACGCTGTCGCGGTT 845
Db 718 CTTAATCGTCAGCAGGCGAAACCTGCTACGAGTCAAGGTCAGAAGGTGCTGAACGTG 777
QY 846 GCGCAACTGAAGATCGGTGGCGCCCAACACTATTCGAACCGCTCGCGCGGTGGCGTG 905
Db 778 AAAGAGATGAAGCTTTTCGCTGACGATACATATACCAATGCGTTAGCGGCGTGGCGTG 837
QY 906 GGCATGCGGTGCGGCTGCGTTCGACGCGATCTCGGCGCTGAAGGCTTTTCGGC 965
Db 838 GCGGATGCGGTAGCTTCGCGCGCGGACGAGTTTGAAGCGGTGAGGATTTTACCGGT 897
QY 966 CTGGCTCATGCTGCCAGTGGGTACGCGAGCGGCGGTGAGCTACTACGAGATTCC 1025
Db 898 CTGGCGATGCTTCCAGCTGGGTTGGAGCATACGCGTTCGTTGGATCAAGACTCG 957
QY 1026 AAGGCACCAACTCGCGCGCGCCCTGGCGGCGATCGAGGGGTGGGTGCCGACATCGAC 1085
Db 958 AAAGCGACCAATTCGCGCAGTACCGAAGCGGCGCTAAACGTTTGCATG-----TGGAC 1011
QY 1086 GGCAGCTGTGCTGCTCGCGCGGAGCGGCAAGGGCGCGATTTCCATGACCTGCGC 1145
Db 1012 GSTACGCTGATCTGCTGGCGCGGAGCGGTAAAGTCGCGCAGACTTTTCTCCGCTGGCG 1071
```

---

```
QY 1146 GAGCGGTCGCGGCTTCTCGCGGGGGTGTACTGCTTGGCGTGTGACGCCGGCTGATT 1205
Db 1072 CGTATCTGTACCGCGGATGTATCCGGCTCTATTGCTTTGGCGGATGGCGGACGCTT 1131
QY 1206 GCCAGGCACTGGGCAACCGGTACCGCTGGTGGCGTGCACAGCTGTGACGAGCAGTC 1265
Db 1132 GCGGCACTG-----GTCGGAAATCGCCCAACAGACTGAGACGATGGAAGAGCGGATG 1185
QY 1266 CGCAGCGCGCGAGCTGGCCCGGAAGGCGATCGGTGCTGTGTTGTCGCGGCTGGCGG 1325
Db 1186 CGTTTCTGGCGCGCACGCTTCAGCCGGGTGATGTTGCTGTGTCGCCCGCTGGCGC 1245
QY 1326 AGCTGTGACATGTTCAAGAACTTCGAAGAACGCGGCGCTGTCGCAAAAGCGGTAGAG 1385
Db 1246 AGCTGTGATCAGTTTAAATAATTTGAGCAACGGGGCGATGCTTTACCCGCTGGCGAAG 1305
QY 1386 GAGCTAGCGTGA 1397
Db 1306 GAGTTAGGTGA 1317
```

RESULT 3  
US-09-741-669-259  
; Sequence 259, Application US/09741669  
; Patent No. US2002022718A1  
; GENERAL INFORMATION:  
; APPLICANT: Forsyth, R. Allyn  
; APPLICANT: Ohlsen, Kari L.  
; APPLICANT: Zyskind, Judith W.  
; TITLE OF INVENTION: Genes identified as required for  
; FILE REFERENCE: ELITRA.009A  
; CURRENT APPLICATION NUMBER: US/09/741.669  
; PRIOR FILING DATE: 2000-12-19  
; PRIOR APPLICATION NUMBER: US 60/173005  
; NUMBER OF SEQ ID NOS: 481  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 259  
; LENGTH: 1317  
; TYPE: DNA  
; ORGANISM: Escherichia coli  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)...(1317)  
US-09-741-669-259

Query Match 19.2%; Score 278.6; DB 9; Length 1317;  
Best Local Similarity 53.7%; Pred. No. 2.6e-56;  
Matches 709; Conservative 0; Mismatches 579; Indels 33; Gaps 5;

```
QY 80 CATCGTTGTCGGGCTCGGCAAGAGCGCATGTCCTGGTGGCTACCTGGCGCGCGCGG 139
Db 27 CGTCAATATCGGCTGGGCTCCACCGGGCTTCTCGTGGACTTTTCTCGCTCGCGG 86
QY 140 CTTGCGCTTTCGCGCTGTCGATACCCGAGAGAACCCCGGAGTGTGCGCCCTCGCGTGC 199
Db 87 TGTGACCGCGCGGTATGATGATGATGATGATGATGATGATGATGATGATGATGATG 145
QY 200 CCAGTATCCGAGGTGGAAGTGTGCGGCGAATCGACCGCGAGTTCCTGCTGCTCGC 259
Db 146 -----AAGCGGTAGAACGCCACACGCGGCGAGTCTGAATGATGATGATGATGATG 197
QY 260 CCGCGAATCTATGTACGCCCCCGGCTGTGCTGCGGACCCCTGCGCTGGTACAGCGCG 319
Db 198 AGATCTGATTTGTCGCCAGTCCCGGTATTTGCACTGGCGCATCTTCTTAAAGCGGTGCG 257
QY 320 CGCGAAGGCGGTGCGGATCTCCGTCGATGATGATGATGATGATGATGATGATGATGATG 379
Db 258 TGATGCGGGAATGAATGCTTGGCGATATGAGCTGTTCTGTGTCGCGAAGACACAGCAC 317
QY 380 GATCGTGGCATCAGCGGTTTCCAAACGCGAAGACCGCTGACCACCCCTGGTGGCGAAT 439
```

```
Db 318 GATTGTGCGATTACCGGTTCTAACGGCAAAAGCACGCTACACGCTAGTGGTGAAT 377
Qy 440 GCGCTGCGCGGCAAGCGTGTGCGCGTGGCGCAACCTCGGCACCGCGGCGCTCGA 499
Db 378 GCGCAAGCGGCGGGTACGTTGGTGGTGGGCAATATTTGGCCCTGCGCTGCTGAT 437
Qy 500 CTGCTGCGCGACGACATCGAGCTGTACGTGTGGAGCTGCGAGCTTCCAGCTGGAAC 559
Db 438 GCTACTGCGATGATGAGTGTGAACGTGTACGTGTGGAAGCTGCGAGCTTCCAGCTGGAAC 497
Qy 560 CTGCGATCGCTCAACGCGGAGTGGCGACCGTGTGAACGTGACGTCAGCGAAGACCATATGA 619
Db 498 CACCTCCAGCTTACAGGCGGTAGCAGCGACCATTTCTGAACGTGACTGAAGATCATATGA 557
Qy 620 TCGCTA---CGACGCGATGCGTACTACCACTGCGCAAGCACGCGATCTCCGCGGTGC 676
Db 558 TCGCTATCCGTTGGTTTACACAGATATCGTGCAGCAAACTGCGCATTTACGAAACGC 617
Qy 677 CCGCAGGTGCTGTGAATCGCGCGATGCGCTGACCGCGCGCTGATCGCGGATACCGT 736
Db 618 GAAAGTTTTCGCTGTTAATGCTGATGATGCGCTTAAACATGCGGATTCGCGTGGGATGA 677
Qy 737 GCGTGTGCTGCTGCTGCGCTGAAACACCGGACTTCAAGCGTTTGGCCCTGATCGAGGA 796
Db 678 ACGTGTGCTGCTGCTTGGCGCTCAACATGGGTGACTATCA-----CCTGAATCATCA 728
Qy 797 AGCGCGCAGAGTGGCTGGCGTTCAGTTTGCAGAGCTGCGCAAGCTGCGCGTTGGCGAAGTAA 856
Db 729 CGAGGCGAAGCTGGCTGCGGTTAAAGGCGAAGAGTGTGAATGTGAAGAGATGA 788
Qy 857 GATCGTGGCGCCCAACTATTCCAGCGCTGCGCGCTGGCGCTGGCGCTGCGCATATCGGT 916
Db 789 ACTTTCGCGCAGCATACTACCAATGCGCTGCGCGCTGGCGCTGGCGATGCTGCG 848
Qy 917 CGGCTGCGCTGCGACCGCTGCTGCGCGCTGAGCGCTTTCGGCCCTGGCTCATCG 976
Db 849 AGGGTTACCGGTCGCGCAGCCCTGAAAGCGCTTAAACCATTTCACTGTGCTGCGCGATCG 908
Qy 977 CTGCGAGTGGTACGCGAGCGGCGTACGCTGAGCTACTACGACGATTCAGAGCGCACCA 1036
Db 909 CTTTGAAGTTGCTGGAGCATACGCGCTAGCTTGGATTACGATTCAAGAGCGACCA 968
Qy 1037 GCTGCGCGCGCTGCGCGGATGCGAGGCTGGGTGCGGACATCGACGCAAGCTGCTGCT 1096
Db 969 GCTGCGCAGTACGGAAGCGGCTGAATGGCTG-----CACGTAGACGCGCACACTGCA 1022
Qy 1097 GCTGCTGCGCGGAGACGCAAGGCGCGGATTTCCATGACCTGCGCGAGCGGTCGC 1156
Db 1023 TTTGTTGCTGGTGGCGATGTAATCGCGCGACTTTAGCCCACTGGCGCGCTTACCTGAA 1082
Qy 1157 GCGCTTTCGCGGCGGTGCTGCTTGGCGCTGACGCGCGGCTGATTGCCCGAGGCACT 1216
Db 1083 TGGCGATAACGTACGTCTGTATTGTTTGGTCTGACGCGCGCGAGCTGGCGGCGCTAC- 1141
Qy 1217 GGGCAACGCGTACGCTGCTGCGGTGCGAAGCGTGGACGAGTTCGGGAGGCGCG 1276
Db 1142 -----GCCCGAAGTGGCAGAACAAACCGAAACTATGGAACAGCGATGCGCTTGTGCG 1196
Qy 1277 CGAGTGGCGCGGAGCGGATGCGTCTGCTGCGCGGCTGCGGAGCGCTGGACAT 1336
Db 1197 TCCGCGTGTTCAGCGGCGGATATGGTTCTCTTCCCGACCTTGTGCCAGCCCTTGATCA 1256
Qy 1337 GTTCAAGAACTTCGAAGACGCGGACGCGCTTTCGCCAAAGCCGTAGAGAGCTAGGCTG 1396
Db 1257 GTTCAAGAACTTTCGAACACAGGCAATGAGTTTGGCGGCTGCTGGCGAAGGATAGGTTG 1316
Qy 1397 A 1397
Db 1317 A 1317
```

RESULT 4

```
US-09-815-242-5932
; Sequence 5932, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5932
; LENGTH: 1317
; TYPE: DNA
; ORGANISM: Escherichia coli
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1317)
US-09-815-242-5932
```

```
Query Match 19.2%; Score 278.6; DB 9; Length 1317;
Best Local Similarity 53.7%; Pred. No. 2.6e-56;
Matches 709; Conservative 0; Mismatches 579; Indels 33; Gaps 5;

Qy 80 CATGTTGTCGCGCTCGGCAAGAGCGCATGCTCCCTGCTGCGCTACCTGGCGCGCGCGG 139
Db 27 CGTCATTATCGCGCTGGGCTCACGGGCTTTCTCGCTGGACTTTTCTCGCTCGCGG 86
Qy 140 CTTGCCCTTTCGCCGTGCTGATACCGAGAACCCGCGGAGCTGGCCACCCCTGCGTGC 199
Db 87 TGTGACGCGCGCGCTTATGGATACGCGTATGACACCGCGCTGGCTGGATAAATACCG- 145
Qy 200 CCAGTATCCGAGGTGGAAGTGGTGGCGGGAAGTGGCGGGAAGTGGCTGCTGCTGCGC 259
Db 146 -----AAGCCGTAGAACCCACACGGGCGAGTCTGAATGATGAATGGCTGATGGCGC 197
Qy 260 CCGCGAACTCTATGTCAGCCCCGCTTGTGCTGCGTGGCACCCCTGCGCTGGTACAGCGCG 319
Db 198 AGATCTGATTTGCGCAGTCCCGGTATTTGCACTGGCGATCCATCTTTAAGCGCTGCGCG 257
Qy 320 CGCGAAAGGCGTGGCGATCTCCGGTGCATGCTCTTTCGCCCGCGAGGCGAAGGCGCC 379
Db 258 TGATGCCGGAATCGAAATCGTTGGCGATATCGAGCTGTTCTGTCGCAAGACACAGCACC 317
Qy 380 GATCGTGGCCATCACCGGTTCCACGCGAAGACACCGTGCACACCCCTGTTGGGCGAAT 439
Db 318 GATTGTGCGGATTACCGGTTCTTAACGCGCAAAAGACGCTACACGCTAGTGGTGAAT 377
Qy 440 GCGCGTGGCGCGGACAAAGCGTGTGCGCGTGGCGGCAACCTCGGCACCGCGCTCGA 499
Db 378 GCGCAAGCGCGGGGTTAACGTTGGTGGTGGCAATATTGGCCTGCCTGCGCTGAT 437
```



```

Db 7361575 GGTGCGGCGCGGACGCTCGAGGAGGCTGCGGCGCGCTTCACGCTGCGCACGCC 7361634
Qy 752 CGGCTGAACAGCGGACTTCAGAGCTTTCGGGCTGATCGAGGAAGCGCCAGAAGTG 811
Db 7361635 CGGACCGTCCAACTCGCGGCTGAGAGGCAATTCGTGCGACCGCCCTTCGTCGAGGA 7361694
Qy 812 GCTGGCGTTCCAGTTCGACAGCTGCTGCGGTTGGGAACTGAAGATCGGTGCGGCCA 871
Db 7361695 CCGCAGAGAACGCCAGGAGCTGCGGAGGTGGCGAGCTGATCCGCCCGCGCGCA 7361754
Qy 872 CAACATATTCAGCGGCTGCGCGGCTGGGCGCATGCGGTGCGGCTGCGGCTGCGGTTGCA 931
Db 7361755 CAACATGCGCAACCGCTTGGCGAGGCGGCTGCGGCGGCTTTCGGGGTGGCGCGCTC 7361814
Qy 932 CGGCATGCTCGCGCGCTGAAGGCTTTTCGGGCTGCTCATCGCTGCGCAATGGGTACG 991
Db 7361815 GCGGTTAGGGACGAGCTGCGGCGCTTCGCGCGGAGCGCCACCGCATGCGCGACGTCGC 7361874
Qy 992 CGAGCGGAGGCGGTGAGCTACTACGACGATTTCCAAAGGCCACCAACGTCGCGCGCGCCCT 1051
Db 7361875 CGAGTGGAGCGGCTACCTACATCGAGAGCTCCAAAGGCCACCAACGCGCGCGGA 7361934
Qy 1052 GCGGCGATCGAGGCGGTGGTGGCGAC-----ATCGAGCGCAAGCTGCTGCTCGC 1105
Db 7361935 AGCCTCTTGGCGGCGTACGCGGTCCATCGTTCGATCGCGGCGGCTCGCCAAAGGCGC 7361994
Qy 1106 CCGCGGAGAGCGCAAGGCGCGGATTTCCATGACCTCGCGGAGCGGCTCGCGGCTTCG 1165
Db 7361995 CTCCTTCGAGAGCTGTTGGCCAGTGGCCAGTGGCACAGCGGCTCGGGGTGTTGCTCATCG 7362054
Qy 1166 CCGGCGGCTGACTGCTTGGCGCTGACGCGCGGCTGATTGCGC---AGGCACTGGGCAA 1222
Db 7362055 TGGGATCGTGGCTGATCGGTGAAGCCCTTGGCGGACAGCGCCCGGAGGTACCCGCTG 7362114
Qy 1223 CGCGTACCGCTGCTGCGCTGCGCAAGCTGGAGAGAGCTCGGCGAGCGCGCGGCT 1282
Db 7362115 CGACCTCGACCGGACGACACTGGGGCGATGCGCGGCTGTCGAGGCGCGGCGGCT 7362174
Qy 1283 GCGCGCGAGGAGATCGGTGCTGTTGTCGCGGCTGCGCGGCTGGCGAGCTGGACATGTTCAA 1342
Db 7362175 CGCGTGGCGGCGACAGCTGCTGCTGCGCGCGGCTGCTGCTGATGACATGTTGCG 7362234
Qy 1343 GAACCTCGAGAGAGCGGCGGCTGTTGCGCAAGCGGTAGAGAGCTAGCGTATGCTG 1402
Db 7362235 CAACATACAAAGCGGCTGACGCGTTTCGCGGAGCGGTTTCGCGGACTCGGAGCCTGATC 7362294
Qy 1403 TCGGTGTTGGCC 1415
Db 7362295 CCGCGCGCTCGCC 7362307

```

RESULT 6

```

US-10-156-761-6097
; Sequence 6097, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109

```

```

; SEQ ID NO 6097
; LENGTH: 1401
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1401)
US-10-156-761-6097

Query Match 12.48; Score 180.2; DB 14; Length 1401;
Best Local Similarity 49.6%; Pred. No. 3.1e-33;
Matches 674; Conservative 0; Mismatches 628; Indels 57; Gaps 6;

Qy 80 CATGTTGTGGCGCTCGGCAAGAGCGCATGTCTCGTGGCTGCTACCTGGCGCGCGCGG 139
Db 27 CACGCTCGCGGGCTCGCGGCTCTCCGGTATCCCGCGCGCGCGCTGCTGACAGCGCTCG 86
Qy 140 CTTGGCTTTTCGCGGTGTCGATACCCGAGAGAACCCCGCGGAGTGCCACCTTCGCTGC 199
Db 87 CGCGGTGCTGACGCTGCTCAACGACGCGGACGAGCGCTCCCGGCGCCAGCGCGGA 146
Qy 200 CCAGTATCCGACAGTGGAGTGGCTTGGGGGAACTCGACGCGGAGTTCCTCTCTCGC 259
Db 147 TCTGGAGGCGCTCGGCATACCGTGCCTCGGTGACGGCGCGACCTGCCCGAGGGAC 206
Qy 260 CCGCGAACTCTATGTCAGCGCGCGCTTGTGCTGCGCACCCCTCGCTGCTGCTGACAGCGC 319
Db 207 CGAGCTCATGTCACACACACCGGCTGCGAGCGGACAAAGCGCTGTTCCCGCGCGCGC 266
Qy 320 CGGAAAGGCGTGGCGATCTCCGGTGACATCGATCTCTTCCCGCGCGAG-----GC 370
Db 267 CGAGGCGGCGCTCCCGCTGCGGCGGAGCTCGAATCGCTGCGGCGGCTGCGCGCGCGC 326
Qy 371 GAAGGCGCGCATGCTGCCATACCGGTTCCAAAGCGAAGAGACCGTGACCAACCTTGGT 430
Db 327 ATCGCGCGCTGCTGCGCTGCGGCGACCAAGCGCAACAGCGACCGTGTGAGATGCT 385
Qy 431 GGGCGAAATGGCGGTGCGCGGACAAAGCTGTGCGCTGCGGCGCAACCTTGCGCACCC 490
Db 387 CGCTCTCATCTGACGCGCGCGGCTGCGCGCGGCTGCGCAACATCGCGCTC 446
Qy 491 GCGCTCGACCTGCTGCGCGCGA-----CATCGAGCTGTAGCTGTGAGCTGTGAG 544
Db 447 GCTCTCGAGCGGCTCTCGCGGAGGAGAGTACGAGCTCTCGCGGTGAACTCTCTCAG 506
Qy 545 CTTCCAGCTGGAACCTTCGATCGCTCAACGCGGAGGTGCGGAGCTGTGTGAACGTGAG 604
Db 507 CTACGAGTGCAGTGGCGCGCTCCCTGCGCGGCGGCGGCGGCGGCTGCTGACATCGC 566
Qy 605 CGAAGACCATATGATGCTGACGAGCGATGCGTACCTACCTTGGCCCAAGCAACCGAT 664
Db 567 GCGCGACCATCGACTGGCACGCGCTCCATGAGGCGCTACACCGCGGCGGCGGCGGCTAT 626
Qy 665 CT-----TCCGCGGTGCGCGCGCGGCTGCTGAGTGAATCGCGCGGATGCGCT 709
Db 627 CTACGAGGCGAATCGGCTGCGCTTACAAAGCTGGCGGCGGCGGCGGCGGAGACCT 686
Qy 710 GACCGCGACCTGATC-----GCGGATACGCTGCGGCTGCTGCTGCTGCT 751
Db 687 GGTGCGCGCGCGGAGCTCGAGGAGGCTGCGCGGCGGCTGCGGCTTACGCTGGGCGCGC 746
Qy 752 CCGCGCTGAACAAGCGGACTTCAAGGCTTTTCGGCGCTGATCGAGGAAGCGGCGGCAAGTG 811
Db 747 CGGACCGTCCAACTCGGCGCTGCTGAGGCGCATTTCTGTCGACCGCGCTTCTGTCGAGGA 806
Qy 812 GCTGGGCTTCCAGTTCGACAGCTGCTGCGGTTGGGAACTGAAGATCGTGGCGCGCA 871
Db 807 CCGCGAAGAACGCCAGGAGCTGCGCGAGGTGGGCGGAGCTCCCGCGCGCGCGCGCA 866
Qy 872 CAACATATTCAGCGGCTCGCGCGGCTGCGGCGGAGCTGCGGCTGCGGCTGCGGCTTCGA 931
Db 867 CAACATCGCAACGCGCTTTCGCGGAGGCGGCGGCTTTCGCGGCTGCGGCGGCTC 926

```







```
;/
;/ NAME/KEY: misc_feature
;/ LOCATION: (36543)..(36543)
;/ OTHER INFORMATION: n equals a, t, g or c
;/ FEATURE:
;/ NAME/KEY: misc_feature
;/ LOCATION: (36551)..(36551)
;/ OTHER INFORMATION: n equals a, t, g or c
;/ FEATURE:
;/ NAME/KEY: misc_feature
;/ LOCATION: (36636)..(36636)
;/ OTHER INFORMATION: n equals a, t, g or c
;/ FEATURE:
;/ NAME/KEY: misc_feature
;/ LOCATION: (40808)..(40810)
;/ OTHER INFORMATION: n equals a, t, g or c
;/ FEATURE:
;/ NAME/KEY: misc_feature
;/ LOCATION: (44416)..(44416)
;/ OTHER INFORMATION: n equals a, t, g or c
;/ FEATURE:
;/ NAME/KEY: misc_feature
;/ LOCATION: (44905)..(44905)
;/ OTHER INFORMATION: n equals a, t, g or c
;/ FEATURE:
;/ NAME/KEY: misc_feature
;/ LOCATION: (44975)..(44975)
;/ OTHER INFORMATION: n equals a, t, g or c
;/ FEATURE:
;/ NAME/KEY: misc_feature
;/ LOCATION: (45593)..(45593)
;/ OTHER INFORMATION: n equals a, t, g or c
;/ FEATURE:
;/ NAME/KEY: misc_feature
;/ LOCATION: (45732)..(45732)
;/ OTHER INFORMATION: n equals a, t, g or c
;/ FEATURE:
;/ NAME/KEY: misc_feature
;/ LOCATION: (47036)..(47036)
;/ OTHER INFORMATION: n equals a, t, g or c
;/ FEATURE:
;/ NAME/KEY: misc_feature
;/ LOCATION: (51334)..(51334)
;/ OTHER INFORMATION: n equals a, t, g or c
;/ FEATURE:
;/ NAME/KEY: misc_feature
;/ LOCATION: (51602)..(51602)
;/ OTHER INFORMATION: n equals a, t, g or c
;/ FEATURE:
;/ NAME/KEY: misc_feature
;/ LOCATION: (51786)..(51786)
;/ OTHER INFORMATION: n equals a, t, g or c
;/ FEATURE:
;/ NAME/KEY: misc_feature
;/ LOCATION: (51805)..(51805)
;/ OTHER INFORMATION: n equals a, t, g or c
;/ FEATURE:
;/ NAME/KEY: misc_feature
;/ LOCATION: (55369)..(55369)
;/ OTHER INFORMATION: n equals a, t, g or c
;/ FEATURE:
;/ NAME/KEY: misc_feature
;/ LOCATION: (65309)..(65309)
;/ OTHER INFORMATION: n equals a, t, g or c
;/ FEATURE:
;/ NAME/KEY: misc_feature
;/ LOCATION: (65313)..(65313)
;/ OTHER INFORMATION: n equals a, t, g or c
;/ FEATURE:
;/ NAME/KEY: misc_feature
;/ LOCATION: (80024)..(80024)
;/ OTHER INFORMATION: n equals a, t, g or c
;/ FEATURE:
```

```
;/
;/ NAME/KEY: misc_feature
;/ LOCATION: (100091)..(100091)
;/ OTHER INFORMATION: n equals a, t, g or c
;/ FEATURE:
;/ NAME/KEY: misc_feature
;/ LOCATION: (102696)..(102696)
;/ OTHER INFORMATION: n equals a, t, g or c
;/ FEATURE:
;/ NAME/KEY: misc_feature
;/ LOCATION: (105121)..(105121)
;/ OTHER INFORMATION: n equals a, t, g or c
;/ FEATURE:
;/ NAME/KEY: misc_feature
;/ LOCATION: (107248)..(107248)
;/ OTHER INFORMATION: n equals a, t, g or c
;/ FEATURE:
;/ NAME/KEY: misc_feature
;/ LOCATION: (117136)..(117136)
;/ OTHER INFORMATION: n equals a, t, g or c
;/ FEATURE:
;/ NAME/KEY: misc_feature
;/ LOCATION: (119750)..(119750)
;/ OTHER INFORMATION: n equals a, t, g or c
;/ FEATURE:
;/ NAME/KEY: misc_feature
;/ LOCATION: (119924)..(119924)
;/ OTHER INFORMATION: n equals a, t, g or c
;/ FEATURE:
;/ NAME/KEY: misc_feature
;/ LOCATION: (120038)..(120038)
;/ OTHER INFORMATION: n equals a, t, g or c
;/ FEATURE:
;/ NAME/KEY: misc_feature
;/ LOCATION: (121344)..(121344)
;/ OTHER INFORMATION: n equals a, t, g or c
;/ FEATURE:
;/ NAME/KEY: misc_feature
;/ LOCATION: (122167)..(122167)
;/ OTHER INFORMATION: n equals a, t, g or c
;/ FEATURE:
;/ NAME/KEY: misc_feature
;/ LOCATION: (122336)..(122336)
;/ OTHER INFORMATION: n equals a, t, g or c
;/ FEATURE:
;/ NAME/KEY: misc_feature
;/ LOCATION: (131340)..(131340)
;/ OTHER INFORMATION: n equals a, t, g or c
;/ FEATURE:
;/ NAME/KEY: misc_feature
;/ LOCATION: (131360)..(131360)
;/ OTHER INFORMATION: n equals a, t, g or c
;/ FEATURE:
;/ NAME/KEY: misc_feature
;/ LOCATION: (139910)..(139910)
;/ OTHER INFORMATION: n equals a, t, g or c
;/ FEATURE:
;/ NAME/KEY: misc_feature
;/ LOCATION: (140398)..(140398)
;/ OTHER INFORMATION: n equals a, t, g or c
;/ FEATURE:
;/ NAME/KEY: misc_feature
;/ LOCATION: (142750)..(142750)
;/ OTHER INFORMATION: n equals a, t, g or c
;/ FEATURE:
;/ NAME/KEY: misc_feature
;/ LOCATION: (145058)..(145058)
;/ OTHER INFORMATION: n equals a, t, g or c
;/ FEATURE:
;/ NAME/KEY: misc_feature
;/ LOCATION: (145171)..(145171)
;/ OTHER INFORMATION: n equals a, t, g or c
;/ FEATURE:
;/ NAME/KEY: misc_feature
```

LOCATION: (145942)..(145942)  
 OTHER INFORMATION: n equals a, t, g or c  
 FEATURE:  
 NAME/KEY: misc\_feature  
 LOCATION: (147197)..(147197)  
 OTHER INFORMATION: n equals a, t, g or c  
 FEATURE:  
 NAME/KEY: misc\_feature  
 LOCATION: (150841)..(150841)  
 OTHER INFORMATION: n equals a, t, g or c  
 FEATURE:  
 NAME/KEY: misc\_feature  
 LOCATION: (152500)  
 OTHER INFORMATION: n equals a, t, g or c  
 FEATURE:  
 NAME/KEY: misc\_feature  
 LOCATION: (152530)..(152530)

Query Match 10.3%; Score 149.8; DB 14; Length 1830121;  
 Best Local Similarity 51.0%; Pred. No. 2.9e-26;  
 Matches 440; Conservative 0; Mismatches 407; Indels 16; Gaps 3;  
 QY 272 TGTACGCGCGCTGTGCTGCGCACCCCTGCGTGTACAGCGCGCGCGAAAGCGT 331  
 DB 1203341 TATAGCCAGCGCTGCGGTAAACACCAAAATTCACACCGCACTTAAAGCGGAGT 1203400  
 QY 332 CGCATCTCCGTCGATCATCTCTCCGCCGAGCGGAGCGCGCGATCGTCGCCAT 391  
 DB 1203401 GGAAGTAATCGCGATATTGAATTATCTGCGCGCACGACAAAGCAATTTGGGGAT 1203460  
 QY 392 CACCGTTCCACGCGAAGACGCGTACACCCCTGTTGGCGGAAATGCGGTGGCGC 451  
 DB 1203461 TACAGTTCAAATGGTAAGTACCGTAACCTTATAGTTATGAATGGCGAAAGTCG 1203520  
 QY 452 GGACAAGCGTTCGCGCTGCGCGCAACCTCGGCACCCGCGCTCGACCTGCTGGCGA 511  
 DB 1203521 TGGTGTGAAGTGTGTATGGCGGAAATATGGGATTCGCCCTTTGTCATTGTAATGA 1203580  
 QY 512 CGACATCGAGTGTACGTGTGGAGCTTCGAGCTTCAGCTGGAACCTGCGATCGCT 571  
 DB 1203581 AGATTGTGAATTTATGTAAGTCTTCTAGTTTTCAGCTTGAGCAACTTATAGTT 1203640  
 QY 572 CAACGCCAGTGGGACCGTGTGACGTGAGCGAAGACCATATGATGCTACGACGG 631  
 DB 1203641 AAAAGCTCGCGACGAGCTGCTTGAAGCTGACTGAAGTATGATGCTATATGGA 1203700  
 QY 632 CATGGCTGACTACCACTGGCCAGCACCGATCTTCGCGGTGCCCGCAGTGTGCT 691  
 DB 1203701 TTTAGAAGATTATCGCCAAAGCAAAATTCAGCATTTATCATTAAGTAGTGTGTT 1203760  
 QY 692 GAATCGCGCGATGCCCTGACCGCTGATCGCCGAT--ACCGTGGCTGCTGCTGTCG 749  
 DB 1203761 GAACAATGAAGTAGGCTGACTTTTGGGAAACGAAATCAAGCAACATACCGTTTC 1203820  
 QY 750 TTCGCGCTGAACAGCGGACTTCAAGCTTTCCGCTGTATCGAGGAAAGCGCGCAAG 809  
 DB 1203821 TTTTCGCGAAATAGTCGGATTATGGCTAAAACT-----GAAATGGCAAGCAA 1203872  
 QY 810 TGGCTGGGTTCCAGTTCGACAAGCTGCTGCGGTGGCGAATCAAGATCCGTGGCGC 869  
 DB 1203873 TATTAATGGTAAAGATGAAGTATTTACCTTGTGAAGAAAGTACATGTTGGTTCG 1203932  
 QY 870 CACAACATTTCAACGCGCTGCGCGCTGGCGTGGCGCATGCGGCTCGCGCTTC 929  
 DB 1203933 CATAATATATGAACATTTTGGCAGCAACAGCATTTGGCACAAGCTATAGTATTAATTA 1203992  
 QY 930 GACCCATGCTCGCGCGCTGAAGCGGTTTTCGCGCTGGCTCATCGCTGCGCATGGGTA 989  
 DB 1203993 GATTCAATTCGTACCGCACTTCGTCAATTTCAAAGGGTTAGATCATCTGTTTCAATAGT 1204052  
 QY 990 CGCGAGCGGAGGCTGAGCTACTACGATTCGAAGGCCACCAAGCTCGCGCGCGC 1049  
 DB 1204053 CATCAAGCTAATGCGATTCGTTGATTAATGACTCTAAGCAACAAATGTGGGAGTACA 1204112

QY 1050 CTGGCGCGATCGAGGGCTGGTCCGACATCGACGCAAGCTGTGCTGCTGCCGCG 1109  
 DB 1204113 GTTGTGCTGATTTGGCTGGCT-----TTATATTAGGGTAAATTGCATTTGTTGCTAGCG 1204166  
 QY 1110 GGAGACGGCAAGGGCGCGCATTT 1132  
 DB 1204167 GGAGACGGAAAGGGCTGATTT 1204189  
 RESULT 10  
 US-09-974-300-1685  
 ; Sequence 1685, Application US/09974300  
 ; Patent No. US20020146721A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Berka, Randy M.  
 ; APPLICANT: Clausen, Ib Groth  
 ; TITLE OF INVENTION: Methods For Monitoring Multiple Gene  
 ; TITLE OF INVENTION: Expression  
 ; FILE REFERENCE: 10085.500-US  
 ; CURRENT APPLICATION NUMBER: US/09/974,300  
 ; CURRENT FILING DATE: 2001-10-05  
 ; PRIOR APPLICATION NUMBER: 09/680,598  
 ; PRIOR FILING DATE: 2000-10-06  
 ; PRIOR APPLICATION NUMBER: 60/279,526  
 ; PRIOR FILING DATE: 2001-03-27  
 ; NUMBER OF SEQ ID NOS: 8481  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 1685  
 ; LENGTH: 1352  
 ; TYPE: DNA  
 ; ORGANISM: Bacillus licheniformis  
 US-09-974-300-1685

Query Match 6.3%; Score 91.8; DB 10; Length 1352;  
 Best Local Similarity 45.7%; Pred. No. 1.7e-12;  
 Matches 490; Conservative 0; Mismatches 562; Indels 21; Gaps 4;  
 QY 327 GCGTGGCCATCTCCGGTGACATCGATCTCTCGCCCGGAGCGCAAGCGCCGATCGTC 386  
 DB 289 GGTATCCGGTCTGGACGGAGTGAACCTCGCTATCATCTGACAGATGCTCCGTTTCATC 348  
 QY 387 GCCATACCGGTTCCAAAGCGAAGACACCGTGACACCCCTGGTGGGGAATGCGGTG 446  
 DB 349 GGCATTACCGTTCCAAAGCGAAGACGACACACGCTTGTATTATGAATGCTGAAG 408  
 QY 447 GCCCGGACACGCTGTCGCGTGGCGGCAACCTCGGCACCCCGCGCTCGACCTG--- 503  
 DB 409 GCCGATTGCGAAAGCGCTCGTTGCTGGGAATATCGGTACGGCGGCAAGCGAGTGCC 468  
 QY 504 ---CTGGCGGACGACATCGAGCTGTACGTGTGGAGCTGTGAGCTTCCAGCTTGAAC 560  
 DB 469 AATCAGCCGCGGAGAGATGATGATGTCGACAGAGTGTGAGTTTTCAGCTCATGGG 528  
 QY 561 TGCATGCTCTCAACGCGGAGTGGCGACCGTGTGTAACGCTGACGAGAGACCATATGAT 620  
 DB 529 ACTTATCAATTCAGACCGCAAGATTGGAATTATTTAAACGTTATTGATGCCATCTCAT 588  
 QY 621 CGCTACGACGCGATGGCTGACTACCACTGCGGCAAGCACCGGATCTCCCGGCTGCCCG 680  
 DB 589 TATCACCATTGCGCGCAAAATATGAGCTTCCCAACACAGAGGCTATCGAAATCAAT 648  
 QY 681 CAGTCTGTGTGAATTCGCGCGATGCGCTGACCGGACCGCTGATCGCGGATACCGTGC 740  
 DB 649 GAAACAGCGTTGCGGTTGCAATTTGGACGATGAGACAGTCTCGCGCTCGCGAGTGT 708  
 QY 741 TGTGTGTGTTGGCGCTGAACACCGGAGCTTCAAGGCTTTCGGCTGTATCGAGAGAC 800  
 DB 709 TCAAAAGCGGAGAGTGTATTTTTCGTCGCGCGGACACTTGAAGCGGAGCGTGGCG 768  
 QY 801 GGCAGAGTGGCTGGCGTTCCTCAGTTCGACAGCTGCTGCGGTTGGGAACTGAAGATC 860  
 DB 769 AAGACGGCGCAATCATGTTTAAACGGCGACCGGCTCATGCGCTGGAAGAGCGTGTGTT 828



; PRIOR APPLICATION NUMBER: JP 2001-272697  
; PRIOR FILING DATE: 2001-08-02  
; NUMBER OF SEQ ID NOS: 15109  
; SEQ ID NO 5987  
; LENGTH: 795  
; TYPE: DNA  
; ORGANISM: Streptomyces avermitilis  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)..(795)  
US-10-156-761-5987

Query Match 4.7%; Score 68.4; DB 14; Length 795;  
Best Local Similarity 44.2%; Pred. No. 5.2e-07;  
Matches 282; Conservative 0; Mismatches 356; Indels 0; Gaps 0;  
QY 53 GAGCTGATCGCTCCGACCACTCCGCATGTTGTCGGCTTCGGCAAGAGCGGCGATGTC 112  
DB 156 GGGCGGGGTACGGCGGACAAACGAGCGGCCCTCGACGGCACCCCGCTCGTCGACGTGGA 215  
QY 113 CTGTGTGCTACCTGGCGCGCGGCTTGCCTTTCGCCGTGGTTCGATACCCGAGAGAA 172  
DB 216 CCGGGCGCAAGATCACTGGCAGCGCGCGGCGAGCTCTCGGGTATCCGATCCAGAA 275  
QY 173 CCGCGCGAGCTGGCCACCCCTGCGTGCAGTATCCGAGGTGAAAGTGCCTTCGGCGGA 232  
DB 276 ACTGCGCGCGGCTCGACGCTGCTGCCACCTCGCGCGCTCGAAGAGGTTCATGATCCG 335  
QY 233 ACTGACGCCGAGTTCCTCTGCTCGCGCGGAACTCTATCTACGCCCGGCTTGTGCT 292  
DB 336 GCTGTGCGGAGTTCGGCTGCGCACCGCGGTGAGGGCGGGCGGGCGGTGCGGT 395  
QY 293 CGCACCCCTCGCTGTGACAGCGCGCGGAAAGCGTGGCGCATCTCCGGTGACATCGA 352  
DB 396 CCGCGGACCCGCTGGAGCAGCGCCAGCCTCGGGGCGCTCTCCCTCGACTTCGACCC 455  
QY 353 TCTCTTCGCGCGGAGGAGCGCGGATGTCGCCATACCGGTTCCAAACGCGAAGAG 412  
DB 456 CCGCTTGGCGGACGACGAGTTCGACCCCGCTCAACGGCGCGGAGTACGCCCGCTCAA 515  
QY 413 CACCGTGACCACTGCTGGCGGAAATGGCGGTGGCGGACAAAGCGTGTGCGCGTGG 472  
DB 516 CGCGGCGGCGCGGAGGACCGCAAGATCTGCGCCATGGGATCCGCTGCGCAAGGG 575  
QY 473 CGGAACCTCGGCAACCGCGGCTGAGCTGCTGGCGGACGACATCGAGCTGTACGTGTT 532  
DB 576 CGTCACCATGACGCTTTCGCCCTGAACTGAACTCGGACACGCTCTCTTCGACAAAGAT 635  
QY 533 GGAGCTGTGAGCTTCCAGCTGGAACCTGCGATCGCTCAACCGCGAGGTGGCGACCGT 592  
DB 636 CATCCGTTGGGGATTCGCGACGCGGCGGTGACGTCTCCCTGACGACTGGGCGCGGA 695  
QY 593 GCTGAAGCTCAGCAAGACCATATGATGCTAGCAGCGGATGCTGACTACCACTGGC 652  
DB 696 GGTACCATCGCGGAGTCTCTCGCGGTTCGGGAGCAAGCAGCTGCGGACGTAAGTGGAGAA 755  
QY 653 CAAGCAGCGATCTTCGCGGTGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 690  
DB 756 CGCGGACCTGAAGCGCGGAGGTGGAACGCGGCGGCGGCGGCGGCGGCGGCGGCGG 793

## RESULT 13

US-10-156-761-1/c  
; Sequence 1, Application US/10156761  
; Publication No. US20030119018A1  
; GENERAL INFORMATION:  
; APPLICANT: OMURA, SATOSHI  
; APPLICANT: IKEDA, HARUO  
; APPLICANT: ISHIKAWA, JUN  
; APPLICANT: HORIKAWA, HIROSHI  
; APPLICANT: SHIBA, TADAYOSHI  
; APPLICANT: SAKAKI, YOSHIYUKI  
; APPLICANT: HATTORI, MASAHARA

; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-262  
; CURRENT APPLICATION NUMBER: US/10/156,761  
; CURRENT FILING DATE: 2002-05-29  
; PRIOR APPLICATION NUMBER: JP 2001-204089  
; PRIOR FILING DATE: 2001-05-30  
; PRIOR APPLICATION NUMBER: JP 2001-272697  
; PRIOR FILING DATE: 2001-08-02  
; NUMBER OF SEQ ID NOS: 15109  
; SEQ ID NO 1  
; LENGTH: 9025608  
; TYPE: DNA  
; ORGANISM: Streptomyces avermitilis  
; FEATURE:  
; NAME/KEY: misc.feature  
; LOCATION: (4187715)  
; OTHER INFORMATION: a, t, c, g, other or unknown  
US-10-156-761-1

Query Match 4.7%; Score 68.4; DB 14; Length 9025608;  
Best Local Similarity 44.2%; Pred. No. 3.3e-07;  
Matches 282; Conservative 0; Mismatches 356; Indels 0; Gaps 0;  
QY 53 GAGCTGATCGCTCCGACCACTTCGCGCATCGTTGTCGGCTTCGGCAAGAGCGGCGATGTC 112  
DB 7237210 GGGCGGGGTACGGCGGACAAACGAGCGGCCCTCGACGGCACCCCGCTCGTCGACGTGGA 7237151  
QY 113 CCGTGTGCGCTACCTGGCGCGCGGCTTGCCTTTCGCCGTGGTTCGATACCCGAGAGAA 172  
DB 7237150 CCGCGGGGCAAGATCACTGGCAGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGG 7237091  
QY 173 CCGCGCGGAGTGGCCACCCCTGCGTGGCCAGTATCCGAGGTGAAAGTGCCTTTCGCGGGA 232  
DB 7237090 ACTCGCGCGCGGCTGCGAGCTGTCGCCACCTCCGCGGCGCTCGAAGAGGTTCATGATCG 7237031  
QY 233 ACTCGAGCGCGAGTTCCTCTGCTCGCGCGCGGAACTCTATGTCAGCCCCCGGCTTGTGCT 292  
DB 7237030 CGTCTGCGGAGTTCGGCGTGGCCACCGCTGAGGGCGGCGGCGGCGGCGGCGGCTG 7236971  
QY 293 CGCGCACCCCTGCGCTGTGACAGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 7236911  
DB 7236970 CCGCGGCGCGGCTGGAGCAGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 7236911  
QY 353 TCTCTTCGCGCGGAGGCGGAGCGCGGATCGTCCGATACCGGTTCCAAACGCGAAGAG 412  
DB 7236910 CCGCTTGGCGGACGAGTTCGACCCCGCTCAACGGCGCGGAGTACGCCCTCCAA 7236851  
QY 413 CACCGTGACCACTGCTGGTGGCGGAAATGGCGGTGGCGGCGGCGGCGGCGGCGGCGGCGG 472  
DB 7236850 CGCGGCGGCGGCGGAGGACCGCAAGATCTGCGGCGGCGGCGGCGGCGGCGGCGGCGG 7236791  
QY 473 CGGCAACCTCGGCAACCCCGCGCTCGACCTGCTGGCGGCGGCGGCGGCGGCGGCGGCGG 532  
DB 7236790 CGTCACCATGCACGGTTCGCCCTGAACTGAACTCGGAGACGCTCTCTTCGACAAAGAT 7236731  
QY 533 GGAGCTGTGAGCTTCAGCTGGAACCTTCGGATGCGCTCAACCGCGGAGGTGGCGACCGT 592  
DB 7236730 CATCCGTTGGGATTCGCGACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 7236671  
QY 593 GCTGAAGCTCAGCGAAGACCATATGATGCTACGAGCGGATGCTGACTACCACTGGC 652  
DB 7236670 GGTACCATCGCGGAGGTCTTGGCGGTCGCGGAGAACCATCTGCGGCGGCGGCGG 7236611  
QY 653 CAAGCAGCGGATCTTCGCGGTGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 690  
DB 7236610 CGCGGACCTGAAGCGCGGCGGAGGTGGAACGCGGCGGCGGCGGCGGCGGCGGCGG 7236573

## RESULT 14

US-09-790-988-1/c  
; Sequence 1, Application US/09790988  
; Patent No. US20020127687A1  
; GENERAL INFORMATION:



**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 14, 2003, 01:43:39 ; Search time 2899 Seconds  
(without alignments)  
12156.430 Million cell updates/sec

Title: US-09-701-229-1

Perfect score: 1450

Sequence: 1 cgtgctgctgcctgcga.....tgtgagcggcagcgcatc 1450

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*

1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estmu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_htc:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_hic:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: em\_gss\_hum:\*  
18: em\_gss\_inv:\*  
19: em\_gss\_pin:\*  
20: em\_gss\_vrt:\*  
21: em\_gss\_fun:\*  
22: em\_gss\_man:\*  
23: em\_gss\_mus:\*  
24: em\_gss\_pro:\*  
25: em\_gss\_rod:\*  
26: em\_gss\_phg:\*  
27: em\_gss\_vrl:\*  
28: gb\_gsl1:\*  
29: gb\_gss2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
c 1	643.6	44.4	892	29	BZ579116 msh2_6205
c 2	638.2	44.0	1445	29	BZ568959 pacs2-164
c 3	382.4	26.4	1171	29	BZ559806 pacs2-164
c 4	232.8	16.1	1194	29	BZ550283 pacs1-60_

c	5	217.2	15.0	928	29	BZ567557	pacs2-164
6	179	12.3	452	10	BE184674	PM4-HT068	
7	69.4	4.8	925	29	CNS0091P	Drosophila	
c 8	67.2	4.6	965	29	CNS070MR	AL423817 T3 end of	
9	66.4	4.6	935	29	CNS006XK	AL066051 Drosophila	
c 10	63.6	4.4	1538	29	AG030607	AG030607 Pan trogl	
c 11	62.6	4.3	925	29	CNS0091P	AL053013 Drosophila	
12	61.6	4.2	583	14	CD43549	CD443549 EL01N0427	
13	61.6	4.2	680	14	CD44231	CD444231 EL01N0437	
14	61.6	4.2	809	14	CD43473	CD434773 EL01N0328	
15	61	4.2	789	14	CD433407	CD433407 EL01N0308	
16	60.6	4.2	844	29	CNS0052P	AL056652 Drosophila	
17	60	4.1	602	14	CD435268	CD435268 EL01N0357	
18	60	4.1	629	14	CD445210	CD445210 EL01N0448	
19	60	4.1	638	14	CD435954	CD435954 EL01N0368	
20	60	4.1	639	14	CD442177	CD442177 EL01N0406	
21	60	4.1	646	14	CD445010	CD445010 EL01N0446	
22	60	4.1	654	14	CD434039	CD434039 EL01N0318	
23	60	4.1	661	14	CD435329	CD435329 EL01N0358	
24	60	4.1	678	14	CD444249	CD444249 EL01N0437	
25	60	4.1	680	14	CD445222	CD445222 EL01N0449	
26	60	4.1	683	14	CD442370	CD442370 EL01N0408	
27	60	4.1	691	14	CD443900	CD443900 EL01N0432	
28	60	4.1	704	14	CD434510	CD434510 EL01N0324	
29	60	4.1	720	14	CD444748	CD444748 EL01N0443	
30	60	4.1	758	14	CD440350	CD440350 EL01N0553	
31	60	4.1	771	14	CD440418	CD440418 EL01N0554	
32	60	4.1	772	14	CD434074	CD434074 EL01N0318	
33	60	4.1	778	14	CD434432	CD434432 EL01N0324	
34	60	4.1	791	14	CD438326	CD438326 EL01N0511	
35	60	4.1	797	14	CD437839	CD437839 EL01N0505	
36	60	4.1	799	14	CD434701	CD434701 EL01N0327	
37	60	4.1	808	14	CD438227	CD438227 EL01N0510	
38	60	4.1	809	14	CD433802	CD433802 EL01N0315	
39	60	4.1	810	14	CD438146	CD438146 EL01N0509	
40	60	4.1	820	14	CD437826	CD437826 EL01N0505	
41	60	4.1	825	14	CD434362	CD434362 EL01N0323	
42	60	4.1	828	14	CD440326	CD440326 EL01N0553	
43	60	4.1	846	14	CD439288	CD439288 EL01N0523	
44	60	4.1	854	14	CD437688	CD437688 EL01N0503	
45	60	4.1	859	14	CD439464	CD439464 EL01N0525	

#### ALIGNMENTS

RESULT 1  
BZ579116/c  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

BZ579116  
msh2\_6205.y2 msh Pseudomonas aeruginosa genomic clone msh2\_6205,  
genomic survey sequence.  
BZ579116  
BZ579116.1 GI:27214177  
GSS.  
Pseudomonas aeruginosa  
Pseudomonas aeruginosa  
Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;  
Pseudomonadaceae; Pseudomonas.  
1 (bases 1 to 892)  
Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M.,  
Burns,J.L., Kaul,R. and Olsen,M.V.  
Whole-Genome-Sequence variation among multiple isolates of  
Pseudomonas aeruginosa library  
J. Bacteriol., (2002) In press  
Contact: Chris K. Raymond  
Genome Center  
University of Washington  
Box 352145, Seattle, WA 98105-2145, USA  
Tel: 2062216954  
Fax: 2066857244  
Email: craymond@u.washington.edu  
Class: shotgun.  
Location/Qualifiers

RESULT	2
BZ568959/c	
LOCUS	
BZ568959	1445 bp DNA linear GSS 17-DEC-2002
pacs2-164_8171.y2	pseudomonas aeruginosa genomic clone
DEFINITION	



```

QY 1283 GGGCCCGAAGGCGATCGGTGCTGTTCGGCGGCGCTCGGAGAGCCTGGACATGTTCAA 1342
Db 523 GGGCCCGAAGGCGATCGGTGCTGTTCGGCGGCGCTCGGAGAGCCTGGACATGTTCAA 1464
QY 1343 GAACTTCGAAGAACGGGAGCGCTGTCGCCAAAGCCGTAGAGGAGCTAGCGTGTGCTG 1402
Db 463 GAACTTCGAAGAACGGGAGCGCTGTCGCCAAAGCCGTAGAGGAGCTAGCGTGTGCTG 404
QY 1403 TCGGTGTTTCGGCCCTTCCTCCGTCGCGCTGTTGAGCCGCGACGGCATC 1450
Db 403 TCGGTGTTTCGGCCCTTCCTCCGTCGCGCTGTTGAGCCGCGACGGCATC 356

RESULT 5
BZ567557/c 928 bp DNA linear GSS 17-DEC-2002
LOCUS pacs2-164_7094.y2 pacs2-164 Pseudomonas aeruginosa genomic clone
DEFINITION pacs2-164_7094, genomic survey sequence.
ACCESSION BZ567557
VERSION BZ567557.1 GI:27199005
KEYWORDS GSS.
SOURCE Pseudomonas aeruginosa
ORGANISM Pseudomonas aeruginosa
REFERENCE Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M.,
AUTHORS Burns,J.L., Kaul,R. and Olsen,M.V.
TITLE Whole-Genome-Sequence variation among multiple isolates of
JOURNAL Pseudomonas aeruginosa library
COMMENT J. Bacteriol., 192 (2002) in press
Contact: Chris K. Raymond
Genome Center
University of Washington
Box 352145, Seattle, WA 98105-2145, USA
Tel: 20622216954
Fax: 2066857244
Email: craymond@u.washington.edu
Class: shotgun.

FEATURES
source
1..928
/organism="Pseudomonas aeruginosa"
/mol_type="genomic DNA"
/strain="2-164"
/db_xref="taxon:287"
/clone="pacs2-164_7094"
/clone_lib="pacs2-164"
/notes="clinical isolate 2-164 whole genomic shotgun
library."

BASE COUNT 238 a 228 c 236 g 226 t
ORIGIN
Query Match 15.0%; Score 217.2; DB 29; Length 928;
Best Local Similarity 71.4%; Pred. No. 2.8e-33;
Matches 327; Conservative 0; Mismatches 128; Indels 3; Gaps 3;

QY 671 CGGTGCGCCCGCAGGTGCTGTTCGGCGCTGAACACCGGACTTCAAGGCTTTCGGCCCTGA 729
Db 539 CGTTGCGCCCGCAGGTGCTGTTCGGCGCTGAACACCGGACTTCAAGGCTTTCGGCCCTGA 480
QY 730 ATACCGTGCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 789
Db 479 ATCCCGTGCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 420
QY 790 TCGAGGAAGACGGCCGAGAGTGGCTTCCAGTTCGACAAAGCTGCTGCGCGGTGGCG 849
Db 419 TCCAAGAGTCTGACAGAGTGGCTGCGGTACGTGTTGACGAATAATTCGGCTGTGCTG 360
QY 850 AACTGAAGATCGGTGGCGCCCAACAATAATTCGAACGCGCTCGCCGCGTGGCGGTGGCC 909
Db 359 ATGTGAAGATCGGTGGCGCCCAACAATAATTCGAACGCGCTCGCCGCGTGGCGGTGGCC 300
QY 910 ATGCGGTGGCGCTGCGCTTCGAC -CCATGCTCGCGCGCTGAAGCGCTTTCGGCGCTG 968

```

```

Db 299 ATGCGGTGCTGCGGTGCTGTTCAGCCATACATGCGGCACCTACAGGTGATCCACGTGT 240
QY 969 GCTATCGCTGCGCAGTGGGTAGCGAGCGCGAGGCGTGTGAG-CTACTAGCAGCATTCCAA 1027
Db 239 TCTCTATCGCTGCGCATGAGTACCTAACACAGGGTCTGTGATGATGTCGAGCAATTACGA 180
QY 1028 GGCCACCAACGTCGGCGCGCCCTCGCGCGCATCGAGGGGCTGGTGGCGACATCGACGCG 1087
Db 179 GCCCGTCTTGGTGGTGCACGCGCTGTTTCGATCGAGTGGATGTTGGTGTGACATTTTCAT 120
QY 1088 CAACTGCTGCTGCTGCTGCGCGCGGAGACGCGAAGGCGC 1125
Db 119 CAACCTCATGCTGCTGCGCAGACGACGACAAAGGGG 82

RESULT 6
BE184674 452 bp mRNA linear EST 22-JUN-2000
LOCUS PM4-HT0688-050500-002-d10 HT0688 Homo sapiens cDNA, mRNA sequence.
DEFINITION BE184674
ACCESSION BE184674
VERSION BE184674.1 GI:8663858
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 452)
Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
20202663
PUBMED 10737800
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?cl=est2=PM4-HT0688-050
500-002-d10&t3=2000-05-05&t4=1)
Seq primer: puc 18 forward
High quality sequence start: 26
High quality sequence stop: 89.
Location/Qualifiers
1..452
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/dev_stage="Adult"
/clone_lib="HT0688"
/notes="Organ: head_neck; Vector: puc18; Site:1: SmaI;
Site:2: SmaI; A mini-library was made by cloning products
derived from ORESTES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the pUC 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."

BASE COUNT 91 a 140 c 133 g 88 t
ORIGIN
Query Match 12.3%; Score 179; DB 10; Length 452;
Best Local Similarity 73.3%; Pred. No. 9.4e-26;

```

BASE COUNT	120 a	61 c	61 g	172 t	511 others
ORIGIN					
Query Match	4.8%;	Score 69.4;	DB 29;	Length 925;	
Best Local Similarity	14.4%;	Pred. No. 0.00058;			
Matches	58;	Conservative 194;	Mismatches 146;	Indels 4;	Gaps 1;
QY	931	ACGCCATGCTCGCGCGCTCAAGCGGTTTTCGGCTGCGCTCATCGCTGCCAGTGGGTAC	990		
Db	524	ANNNNAGCSMCGKCGSTTBGTTTTTTSSGSGYKGCSSGGBSCSCSCSCSS 583			
QY	991	CGCAGCGGCGAGGCGGTGAGCTACTACGACGATTCCAAGCGACCAACGTGCGGCGCGCCC	1050		
Db	584	SCSCCBBCCCSCSSYCCSSBSBSSKCSFBSBSCSCCCKSVCGTSCSSSSCSBSSS 643			
QY	1051	TGCGGGGATCGAGGGCTGGTCCGACATCAGCAGGCAAGCTGTGTCTGCCCGCG 1110			
Db	644	TSSSTTSSTTSKSSSGSSSSSSSTTSKSTASGSGWSAGGGSGTGSTSSSSSSS 703			
QY	1111	GAGACGCCAAGCGCGCGATTCATGACCTGCGCGAGCGCGTTCGTGCGGG 1170			
Db	704	TSTSSSSVSGSKSTBSSGSSSSSSSTSSBBSCTSTSSSSSYSTSCCTC 763			
QY	1171	CGTGTGTACTGCTTGCCCTGACCGGGCTGATGCCCAGGCACTGGGCAACGCGGTAC 1230			
Db	764	CCSY----SYSSSTSSSSTSWGTSKSSSSSVGTSSSSDSTTCCSCCCYMTCCSTYB 819			
QY	1231	CGCTGTGTCGGCTCGCAACGCTGACCAAGCAGTCCGGCAGGCGCGCTGCGCGCG 1290			
Db	820	MBCTYSTSCGSSSSSGKGVTKCGCGCGSSSTNGMBGTSSACSSSSSCSSSSVSSS 879			
QY	1291	AAGCGATGCGGTGCTGCTTTCGCGCGCCCTGCGGAGCCTGG 1332			
Db	880	KSASSSSSVSSGSGSVSSNSSSAKSSSSGSGVSSSGSGSG 921			

840 MBUITSICGGSSSSSGAGGVTKGCGCCGSSSINGMEGTSSACSSSSSSSCSSSSSVSSSS 819

DQ

QY 1291 AAGCGATGGTGCTTTTCGCCGCCCTCGCGAGCCTGG 1332

DB 880 KSAASSSVSSGSGVSSNNSSASKSSSGSVSSGSGSG 921

RESULT 8	CNS070MR	965 bp	linear	GSS 06-JUL-2001
LOCUS	CNS070MR/c			
DEFINITION	T3 end of clone AZ0AA010F07 of library AZ0AA from strain CBS 712 of Kluyveromyces marxianus, genomic survey sequence.			
ACCESSION	AL423817			
VERSION	AL423817.1			
KEYWORDS	GSS.			
SOURCE	Kluyveromyces marxianus			
ORGANISM	Kluyveromyces marxianus			

ORGANISM  
Kluyveromyces fragilis  
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
Saccharomycetales; Saccharomycetaceae; Kluyveromyces.  
REFERENCE  
1 (bases 1 to 965)  
AUTHORS  
Souclet J.-L., Aigle, M., Artiguenave, F., Blandin, G.,  
Bolotin-Fukuhara, M., Bon, E., Bottier, P., Casaregola, S.,  
de-Montigny, J., Dujon, B., Durrens, A., Lepingle, B.,  
Malpertuy, A., Neuvéglise, C., Ozler-Kalogeropoulos, O., Potier, S.,  
Saurin, W., Tekalia, F., Toffano-Nioche, C., Wesolowski-Louvel, M.,  
Wincker, P. and Weissenbach, J.  
TITLE  
Genomic exploration of the hemiascomycetous yeasts: 1. A set of  
yeast species for molecular evolution studies  
JOURNAL  
FEBS Lett. 487 (1), 3-12 (2000)

JOURNAL MEDLINE  
PUBMED  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
Submitted (08-SEP-2000) Genoscope - Centre National de Sequencage,  
Kruyeromyces marxianus var. marxianus  
FEMS Lett. 487 (1), 71-75 (2000)  
20584722  
3 (bases 1 to 965)  
11152887  
Genoscope.  
Direct Submission

2 rue Gaston Cremieux, Cp 5706, 91057 EVRY cedex, FRANCE. (E-mail : segref@genoscope.cns.fr - Web : www.genoscope.cns.fr)  
This GSS is part of a random genomic sequencing program of thirteen yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii, Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila, Candida tropicalis and Yarrowia lipolytica. Genomic Inserts of 3 to 5 kb were prepared and both extremities were sequenced. See keywords for description of this sequence and for the sequence of the other extremity of this insert.

COMMENT

FEATURES

source

1..965

/organism="Kluyveromyces marxianus"

/mol\_type="genomic DNA"

/strain="CBS 712"

/variety="marxianus"

/db\_xref="taxon:4911"

/clone="AZ0AA010F07"

/clone\_lib="AZ0AA"

/note="end : T3"

complement(<355..>836)

/note="similar to CAB08672 [ murd ]

1 putative frameshift(s)"

/evidence=not\_experimental

BASE COUNT 199 a 275 c 310 g 159 t 22 others

ORIGIN

Query Match 4.6%; Score 67.2; DB 29; Length 965;

Best Local Similarity 47.6%; Pred. No. 0.0016;

Matches 226; Conservative 11; Mismatches 230; Indels 8; Gaps 2;

QY 916 TCGGCTCGCTTCACGCCATCTCGCGCGCTCAAGCGCTTTCCGGCTGGCTCATC 975

DB 838 TCGGCTCGCGCGCGCGCTGGTCAAGGCTCTCGCGACTCCACCTTSGCGCATC 779

QY 976 GCTGCCATGGTACGCGAGCGGCGGCTGAGTACTACGAGATTCCAAGGCCACA 1035

DB 778 GTATCGAGACCGTCCATAAAGCGGTGGCATCACSKGGKGGAGACTCCAAGGCCACW 719

QY 1036 ACGTCGCGCGCTCGCGCATCGAGGGCTGGTGCCGAC-----ATCCAGCGCA 1089

DB 718 ATYTTTTTTTGGCTTCYCTCTGTCGCGCTTCGAGACATCTTGTGATGTCGCGGAG 659

QY 1090 AGCTGGTCTGCTCGCGCGGAGACGCGCGCGATTCATGACCTGGCGGAGC 1149

DB 658 GCCAGGCCAAGGCCACCTTGTGAGACCTGGTCAACCCACGCCGAGAGTGGCGG 599

QY 1150 CGTTCGCGCTTTCGCGCGGCTG--GTACTGCTTGGCGTGCAGCGCGGCTGATTC 1207

DB 598 AGCCATCGTGTGGGAACCGACCTGACGTGCTGCCAAGTCNTTGGCGGACGCGCC 539

QY 1208 CCAGSACTGGGCAACGGGTACCGTGTGGCGCTGCCAAGCTGGAGCAAGCATCCG 1267

DB 538 GCATGTCCCGCTGCTGATCATGATCTTTCGCGGAGATGCGGCAAGCGATTCG 479

QY 1268 GCAGCGCGGAGCTGGCGCGGAGCGATCGGTGCTGTCGCGCGCTCGCGGAG 1327

DB 478 CGAGCTCCGAGATGGCGCGCGCGGACACCGTCTTGATGCTCGGGGTGGCGAG 419

QY 1328 CTTGACATGTTCAAGAACTTCGAAGACGGGAGCGCTGTTCCGCAAGCCGTA 1382

DB 418 CTTGAAATTTGGAATGATATGATMMMGCGGTGACGATTTCTGTCAGCGCGCA 364

RESULT 9

CNS006XK

LOCUS

DEFINITION

Drosophila melanogaster genome survey sequence T7 end of BAC # BACR14N09 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.

ACCESSION

AL06051

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Location/Qualifiers

1..935

/organism="Drosophila melanogaster"

/mol\_type="genomic DNA"

/db\_xref="taxon:7227"

/clone="BACR14N09"

/clone\_lib="RPCI-98"

/note="end : 17"

BASE COUNT 257 a 170 c 162 g 96 t 250 others

ORIGIN

Query Match 4.6%; Score 66.4; DB 29; Length 935;

Best Local Similarity 29.9%; Pred. No. 0.0023;

Matches 123; Conservative 108; Mismatches 180; Indels 1; Gaps 1;

QY 923 GCCGTTTCAGCCATGCTCGCGCGCTGAAGCGCTTTCCGGCTGCGCTCATCGTCCA 982

DB 496 SSTGTMDMMYCGCGSCMCTSSGTCGCTCCMYMSSVSCCSCGCGCSCCTSK 555

QY 983 GTGGGTACGCGAGCGGCGGTGAGCTACTACGACGATTCCAAGGCCACCACTCGG 1042

DB 556 MCSTYGCCKCGCGCTSCSSSSCCSBSYSTCCTBCTKCSGCTGSGCTGCCGGG 615

QY 1043 CGCCGCTTCGCGCGATCGAGGGCTGGGTGCGGACATCGACGCAAGCTGTGCTCT 1102

DB 616 SGCG 675

QY 1103 CGCGCGCGGAGACGGCAAGCGCGCGCTTTCATGACGCTGCGGAGCGGTGCGCGCT 1162

DB 676 SGCG 735

QY 1163 CTGCGCGCGGTGCTGCTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1222

DB 736 CSCCG 795

QY 1223 CGCGCTACCGCTGCTGCGCGCTGCGCAAGCTGCGAGACGAGTCCGCGCGCGCGCG 1282

DB 796 SGCG 854

QY 1283 GGCCCCCGAAGCGCATCGGTGCTGTTGTCGCCGCGCTGCGCGAGCTGGAC 1334

DB 855 SGSSCCG 906

RESULT 10

AL06051.1 GI:4945019

GSS.

Drosophila melanogaster (fruit fly)

Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

1 (bases 1 to 935)

Genoscope.

Direct Submission

Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :

BP 191 91006 EVRY cedex - FRANCE (E-mail : segref@genoscope.cns.fr

- Web : www.genoscope.cns.fr)

Determination of this BAC-end sequence was carried out as part of a

collaboration with the Berkeley Drosophila Genome Project (BDGP).

The BDGP is constructing a physical map of the Drosophila

melanogaster genome using these BACs. For further information

please see <http://www.fruitfly.org> The BDGP Drosophila

melanogaster BAC library was prepared by Kazutoyo Osoegawa and

Aaron Mammoser in Pieter de Jong's laboratory in the Department of

Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,

NY. The library is named RPCI-98 and was constructed by partial

EcoRI digestion of Drosophila DNA provided by the BDGP from the

isogenic strain y2; cn bw sp, the same strain used for the BDGP's

P1 and EST libraries. A more detailed description of the library

and how to order individual BAC clones, the entire library, or

filters for hybridization from the BACPAC Resource Center can be

found at [http://bacpac.med.buffalo.edu/drosophila\\_bac.htm](http://bacpac.med.buffalo.edu/drosophila_bac.htm).

Location/Qualifiers

1..935

/organism="Drosophila melanogaster"

/mol\_type="genomic DNA"

/db\_xref="taxon:7227"

/clone="BACR14N09"

/clone\_lib="RPCI-98"

/note="end : 17"

```

AG030607/c
LOCUS          1538 bp      DNA      linear      GSS 01-NOV-2001
DEFINITION    Pan troglodytes DNA, clone: PTB-003A18.F, genomic survey sequence.
ACCESSION     AG030607
VERSION       AG030607.1 GI:16557480
KEYWORDS      GSS.
SOURCE        Pan troglodytes (chimpanzee)
ORGANISM      Pan troglodytes
REFERENCE     1
AUTHORS       Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
               Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE         BAC end sequences of Library PTB
JOURNAL       Unpublished
REFERENCE     2 (bases 1 to 1538)
AUTHORS       Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
               Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE         Direct Submission
JOURNAL       Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
               and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
               1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
               (E-mail:chimpbes@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
               Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT       Clones are derived from the chimpanzee BAC library PTB This BAC end
               was generated during the R&D process and may have higher chance of
               clone tracking errors.
PRIMERS
Sequencing: -21M13
LIBRARY
Vector        : PKS145
R.Site 1      : SacI
R.Site 2      : SacI.
Location/Qualifiers
1. 1538
   /organism="Pan troglodytes"
   /mol_type="genomic DNA"
   /db_xref="taxon:9598"
   /clone="PTB-003A18.F"
   /sex="male"
   /cell_type="lymphoblast"
   /clone_lib="PTB Chimpanzee Male BAC Library"
   /size=669 c 552 g 64 t 186 others

BASE COUNT    67 a 669 c
ORIGIN
Query Match   4.4%; Score 63.6; DB 29; Length 1538;
Best Local Similarity 40.6%; Pred. No. 0.0089;
Matches 274; Conservative 0; Mismatches 390; Indels 11; Gaps 2;

QY 626 CGACGGCATGGCTGACTACACCTGGCCAAAGCAGCCGATCTCCGCGGTGCCGCCAGGT 685
DB 1381 CGNCNGCGNNGGNCNGCNGNCGNGTGTGNNNGCGCGCGTTTGC GCGGNGCNCNGN 1322

QY 686 CGTGGTGAATCGCGCCATGCCCTGACCCAGCCGCTGATCGCGATACCGTGGCGTGTG 745
DB 1321 GGGGGGNNCGCGCGCGGGGNGCGCGGNGCGCGGNGCGCGGNGCGCGGCGCGGCGT 1262

QY 746 GTCTTCGGCTGAACAAGCCGAGCTTCAAGGCTTTCGGGCTGATCGAGGAAGACGCCA 805
DB 1261 CCGTTTCGGNNGCGCGGCGCGCGCGGNGCGCGGCGCGCGGCGCGGCGCGCGCG 1209

QY 806 GAAGTGTGCTGCCTTCAGTTCACAGCTGCTGCCGTTTGGCGAATGAAGATCCGTGG 865
DB 1208 GGAGCGCGCCNCGCGCGCGCGCGGNGCGGNGCGCGCGCGGNGCNGCGGCGCGC 1149

QY 866 CGCCACAACTATTCCAAACGGCTGCGCGCGCTGGCGCATCGGTCGCGCTGCC 925
DB 1148 CGCNGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1089

QY 926 GTTCGAGCCATGCTCGCGCGCTGAAGCGCTTTCGGGCTGGCTCATGCTGCCAGTG 985
DB 1088 GNGNCCGCGCGGNGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1029

```

```

QY 986 GTAGCGGAGCGGCGGCGGTGAGCTACTACG-----ACGATTCCAAAGCCACCAACGTCG 1041
DB 1028 CGNGCGCGNCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 969

QY 1042 GCGCCGCCCTGGCGGCGATCGAGGGGCTGGTGGCGACATCGACGGCAAGCTGGTGTGC 1101
DB 968 NGGCGCGNCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 909

QY 1102 TCGCGCGCGGAGAGCGGCAAGGCGCGCGCATTTCCATGACCTGCGGAGCGCGTCCGCGCT 1161
DB 908 NCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 849

QY 1162 TCTGCCGGGCGGTGTACTGCTGTGGCGGTGACGCCGCGGTGATGCCCCAGGCACTGGCA 1221
DB 848 GCGCGCGCGGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 789

QY 1222 ACCGGGTACCGCTGCTGCGGTGCGCAACGCTGGAGAGCAGTCCGCGAGCGCGCGCG 1281
DB 788 GCNCGNCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 729

QY 1282 TGGCCCCCGAAGGCG 1296
DB 728 GCGCNGCGGCGCGCG 714

RESULT 11
CNS0091P/c
LOCUS          925 bp      DNA      linear      GSS 03-JUN-1999
DEFINITION    Drosophila melanogaster genome survey sequence TET3 end of BAC #
               BAC19D16 of RPCI-98 library from Drosophila melanogaster (fruit
               fly), genomic survey sequence.
ACCESSION     AL053013
VERSION       AL053013.1 GI:4934461
KEYWORDS      GSS.
SOURCE        Drosophila melanogaster (fruit fly)
ORGANISM      Drosophila melanogaster
               Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota;
               Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
               Ephydroidea; Drosophilidae; Drosophila.
REFERENCE     1 (bases 1 to 925)
AUTHORS       Genoscope.
TITLE         Direct Submission
JOURNAL       Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
               BP 191 91006 Evry cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
               - Web : www.genoscope.cns.fr)
COMMENT       Determination of this BAC-end sequence was carried out as part of a
               collaboration with the Berkeley Drosophila Genome Project (BDGP).
               The BDGP is constructing a physical map of the Drosophila
               melanogaster genome using these BACs. For further information
               please see http://www.fruitfly.org The BDGP Drosophila
               melanogaster BAC library was prepared by Kazutoyo Osogawa and
               Aaron Mannosier in Pieter de Jong's laboratory in the Department of
               Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
               NY. The library is named RPCI-98 and was constructed by partial
               EcoRI digestion of Drosophila DNA provided by the BDGP from the
               isogenic strain y2; cn bw sp, the same strain used for the BDGP's
               p1 and EST libraries. A more detailed description of the library
               and how to order individual BAC clones, the entire library, or
               filters for hybridization from the BACPAC Resource Center can be
               found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES
   source      1. 925
               /organism="Drosophila melanogaster"
               /mol_type="genomic DNA"
               /db_xref="taxon:7227"
               /clone="BAC19D16"
               /clone_lib="RPCI-98"
               /note="end : TET3"

BASE COUNT    120 a 61 c 61 g 172 t 511 others
ORIGIN
Query Match   4.3%; Score 62.6; DB 29; Length 925;
Best Local Similarity 13.4%; Pred. No. 0.013;

```

[illegible]

RESULT 12  
CD443549  
LOCUS CD443549 583 bp mRNA linear EST 03-JUN-2003  
DEFINITION EL01N0427H12.b Endosperm\_4 Zea mays cDNA, mRNA sequence.

SEI.	Zea mays	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Zea.
REFERENCE	1 (bases 1 to 583)	
AUTHORS	Lai, J., Dey, N., Kim, C.S., Becraft, P., Larkins, B., Linton, E. and Messing, J.	
TITLE	Sequencing of the maize endosperm ESTs	
JOURNAL	Unpublished	
COMMENT	Contact: Lai, Jinsheng Dr. Joachim Messing's lab Waksman Institute, Rutgers University 190 Frelinghuysen Rd., Piscataway, NJ 08854, USA Tel: 732-445-3801 Fax: 732-445-5735 Email: jlai@waksman.rutgers.edu Seq primer: T3.	

```

seq primer: 13:
Location/Qualifiers
1. .583
/organism="Zea mays"
/mol_type="mRNA"
/cultivar="W22"
/db_xref="taxon:4577"
/tissue_type="Endosperm of 7-23DAP"
/clone_lib="Endosperm_4"
/note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
XhoI"
69 a 181 c 243 g 90 t
BASE COUNT
ORIGIN

```

Query Match	4.2%	Score 61.6;	DB 14;	Length 583;
Best Local Similarity	46.3%	Pred. No. 0.019;		
Matches 202; Conservative	0;	Mismatches 234;	Indels 0;	Gaps 0;

Qy	884	CGCGCTCGCCGCGCTGGCGCTGGGCCATCGGCTCGCCTGCCGTTCGACGCCATGCTCGG	943
Db	147	CGCGGGCGTGGTCGAGAGAGTGGCGCGCGCGCTGACGACCTGTCTCGTGGCGACCGCGT	206
Qy	944	CGCGCTGAAGGCCGTTTTTCGGCGTGGCTCATCGCTGCCAGTGGGTACGGAGCGGCAGGG	1003
Db	207	GGCGCTGGAGCCGGGCGTCAGCTGCTGGCGCTGCCGCCACTGCAAGGGCGGCGGTACAA	266
Qy	1004	CGTGAGCTACTACGAGGATTCCAAGGCCACCAACAGTTCGGCGCGCCCTCGCGCCCATCGA	1063
Db	267	CTGTGCGAGGACATGAAGTTCTTCGCCACCCCGCCGGTGCAGGCTCCTGCGCAACCA	326
Qy	1064	GGGCGTGGTGCCGACATCGACGGCAAGCTGGTGTCTGCCGGGGGAGAGCGCAAGGG	1123
Db	327	GGTGGTGCACCCCGGCGACCTGTGTCTCAAGCTCCCGGACGGGTGAGCTCGAGGAGGG	386
Qy	1124	CGCCGATTTCCATGACTGCGCGAGCGGTGCGGCGCTCTCGCGGGCGGTGTACTGCT	1183
Db	387	CGCCATGTGCGAGCCGCTGAGCGTGGGCGTGCACGCGTCCCGCGCGGGGGTGGGGCC	446
Qy	1184	TGGCGGTGACCGCGGCGCTGATTTGCCACGAGACTTGGGCACGCGGTACCGCTGGTGGCGGT	1243
Db	447	CGAGAGGGGCGTCTCGTGTGGGCGCCGCCCATCGGCCCTGGTGTCTGCTGCGCGCGC	506
Qy	1244	CGCAACGCTGGAGGAAGCAGTCCGGCAGGCGCGGAGCTGGCCCGGAAGGGCGATCGCGT	1303
Db	507	CGCGGCGTTCGGGCGCCCGGCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT	566
Qy	1304	GTGTGTGTGCGCGGCC	1319
Db	567	GGCCAGGTGCTGGGC	582

RESULT	13
CD444231	
LOCUS	CD444231
DEFINITION	680 bp mRNA linear EST 03-JUN-2003
	ELO1N0437E03.b Zea mays cDNA, mRNA sequence.

CD444231	GI:31359874
CD444231.1	
EST.	
KEYWORDS	
SOURCE	Zea mays
ORGANISM	Zea mays
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Zea.
AUTHORS	Lai,J., Dey,N., Kim,C.S., Becraft,P., Larkins,B., Linton,E. and Messing,J.
TITLE	Sequencing of the maize endosperm ESTs

TITLE  
 JOURNAL  
 COMMENT  
 Sequencing of the maize endosperm ESTs  
 Unpublished  
 Contact: Lai, Jinsheng  
 Dr. Joachim Messing's lab  
 Waksman Institute, Rutgers University  
 190 Frelinghuysen Rd., Piscataway, NJ 08854, USA  
 Tel: 732-445-3801  
 Fax: 732-445-5735  
 Email: jlai@waksman.rutgers.edu  
 Seq primer: T3.

```

seq primer: 13:
FEATURES
    source            Location/Qualifiers
    1..680
        /organism="Zea mays"
        /mol_type="mRNA"
        /cultivar="W22"
        /db_xref="taxon:4577"
        /tissue_type="Endosperm of 7-23DAP"
        /clone_lib="Endosperm_4"
        /note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
        XhoI"

```

```

BASE COUNT      85 a      207 c      288 g      99 t      1 others
ORIGIN

Query Match      4.28; Score 61.6; DB 14; Length 680;
Best Local Similarity 46.3%; Pred. No. 0.019;

```





Query Match	4.2%	Score 61;	DB 14;	Length 789;
Best Local Similarity	46.6%	Pred. No. 0.026;		
Matches 196;	Conservative 0;	Mismatches 225;	Indels 0;	Gaps 0;

  

QY	899	GGCGCTGGGCATGCGCTCGCGCTGCCGCTTCGACGCGCATGCTCGGCGCGCTGAAGCGCTT	958	
Db	23	GGAGGTGGGGCGCGCGTGACCTTCTGTCCGTGGCGGACCGCGTGGCGCTGAGGCGCGG	82	
QY	959	TTCCGGCGCTGGCTCATCGCTGGCAGTGGGTACCGAGCGCGCAGGGCGGTGAGCTACTACGA	1018	
Db	83	CGTCAGCTGTGGCGCTGCGCGCTGCGCACATGCAAGGCGGGCGGTACAACCTGTGCGAGGACAT	142	
QY	1019	CGATTTCCAAAGGCCAACCAAGCTCGGCGCGCCCTTGGCGCGCATGCGAGGGCGTGGTGCCGA	1078	
Db	143	GAAGTTCTTCGCACACCCCGCGGTGCACGGCTGCTGGCGGACCAAGGTGGTGACCCCGCG	202	
QY	1079	CATTCGACGGCAAGCTGGTGTCTGCTCCCGCGCGAGACGGCAAGGGCGCGGATTTCATGA	1138	
Db	203	CGACCTGTGTCTTCAAGCTCCCGACGGGTGACCTTGGAGAGGGCGCCATGTGCGAGCC	262	
QY	1139	CTTGCGCGAGCGGTGCGCGCTTTCGCGCGGGGTGCTACTCTTTGGCGGTGACGCCGG	1198	
Db	263	GCTGACGCTGGCGCTGCACGCGTGCCTCCCGCGGGGGTGGGCGCCGAGACGGCGCTGCT	322	
QY	1199	GCTGATTGCCCGGCACTGGGCAACCGGTACCGCTGGTGCCTGTCGCAACGCTGGACGA	1258	
Db	323	CGTGGTGGCGCGCGGCCCATCGGCCCTGGTGTGCTGCTGGCGCGCGGGCCCTTCGCGCG	382	
QY	1259	AGCAGTCCGCGAGGGCGCGCGAGCTGGCCCGCGAAGCGGATGCGGTGCTGTGTGCGCGCG	1318	
Db	383	GGCGCGCTGTGGTCTGGATGTGGACGACCACCGCCCTGGCGCGTGGCGAGGTGCGTGGG	442	
QY	1319	C	1319	
		+		
Db	443	C	443	

Search completed: August 14, 2003, 04:12:24  
Job time : 2903 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model  
Run on: August 14, 2003, 00:43:39 ; Search time 422 Seconds  
(without alignments)  
9275.320 Million cell updates/sec

Title: US-09-701-229-1  
Perfect score: 1450  
Sequence: 1 cgtgctgctgctgctgcca.....tggtgacgcgcgcgcgc 1450

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0  
Searched: 2552756 seqs, 1349719017 residues  
Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N\_Geneseq\_19Jun03.\*  
1: /SIDS1/gcgdata/geneseq/geneseq-nbml/NA1980.DAT.\*  
2: /SIDS1/gcgdata/geneseq/geneseq-nbml/NA1981.DAT.\*  
3: /SIDS1/gcgdata/geneseq/geneseq-nbml/NA1982.DAT.\*  
4: /SIDS1/gcgdata/geneseq/geneseq-nbml/NA1983.DAT.\*  
5: /SIDS1/gcgdata/geneseq/geneseq-nbml/NA1984.DAT.\*  
6: /SIDS1/gcgdata/geneseq/geneseq-nbml/NA1985.DAT.\*  
7: /SIDS1/gcgdata/geneseq/geneseq-nbml/NA1986.DAT.\*  
8: /SIDS1/gcgdata/geneseq/geneseq-nbml/NA1987.DAT.\*  
9: /SIDS1/gcgdata/geneseq/geneseq-nbml/NA1988.DAT.\*  
10: /SIDS1/gcgdata/geneseq/geneseq-nbml/NA1989.DAT.\*  
11: /SIDS1/gcgdata/geneseq/geneseq-nbml/NA1990.DAT.\*  
12: /SIDS1/gcgdata/geneseq/geneseq-nbml/NA1991.DAT.\*  
13: /SIDS1/gcgdata/geneseq/geneseq-nbml/NA1992.DAT.\*  
14: /SIDS1/gcgdata/geneseq/geneseq-nbml/NA1993.DAT.\*  
15: /SIDS1/gcgdata/geneseq/geneseq-nbml/NA1994.DAT.\*  
16: /SIDS1/gcgdata/geneseq/geneseq-nbml/NA1995.DAT.\*  
17: /SIDS1/gcgdata/geneseq/geneseq-nbml/NA1996.DAT.\*  
18: /SIDS1/gcgdata/geneseq/geneseq-nbml/NA1997.DAT.\*  
19: /SIDS1/gcgdata/geneseq/geneseq-nbml/NA1998.DAT.\*  
20: /SIDS1/gcgdata/geneseq/geneseq-nbml/NA1999.DAT.\*  
21: /SIDS1/gcgdata/geneseq/geneseq-nbml/NA2000.DAT.\*  
22: /SIDS1/gcgdata/geneseq/geneseq-nbml/NA2001A.DAT.\*  
23: /SIDS1/gcgdata/geneseq/geneseq-nbml/NA2001B.DAT.\*  
24: /SIDS1/gcgdata/geneseq/geneseq-nbml/NA2002.DAT.\*  
25: /SIDS1/gcgdata/geneseq/geneseq-nbml/NA2003.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	1450	100.0	1450	21	AAZ47131
2	1347	92.9	1347	23	AAS34275
3	332.8	23.0	1317	23	AAS56341
4	278.6	19.2	1317	22	AAH81460
5	278.6	19.2	1317	23	AAS52295
6	180	12.4	1380	21	AAZ53175
7	180	12.4	69936	21	AAH81479
8	180	12.4	349980	21	AAF21607

9	180	12.4	1437668	21	AAA81490
10	173.8	12.0	1461	22	AAH52025
c 11	173.8	12.0	4411529	22	AAJ99682
c 12	172.2	11.9	4403765	22	AAJ99683
13	165.6	11.4	1335	25	ABZ41528
14	158	10.9	4942	23	AAH89816
15	158	10.9	4944	23	AAS89273
16	158	10.9	4944	23	AAS90181
17	157.2	10.8	1080	23	AAS89258
18	157.2	10.8	1080	23	AAS89801
19	157.2	10.8	1080	23	AAS90166
20	152	10.5	2058	23	AAS73012
21	152	10.5	4831	23	AAS94542
22	151.4	10.4	3117	23	AAS69095
23	151.4	10.4	3117	23	AAS89240
24	149.8	10.3	1314	23	AAS53453
25	149.8	10.3	1830121	17	AAJ42063
c 26	119	8.2	96109	22	AAF28548
27	110	7.6	2703	23	AAS64579
28	110	7.6	2703	23	AAS66661
29	91.8	6.3	1352	24	ABK74394
c 30	86	5.9	34980	22	AAH68531
c 31	79.8	5.5	356	21	AAZ53173
c 32	78.4	5.4	356	21	AAZ53174
33	76	5.2	9515	15	AAQ55145
34	76	5.2	9515	25	ABZ77356
35	76	5.2	9515	25	AAJ51844
c 36	74.6	5.1	37856	21	AAJ11992
c 37	73.6	5.1	356	21	AAZ53172
38	73.6	5.1	44377	18	AAZ78508
39	73.6	5.1	44377	18	AAJ80414
c 40	73.2	5.0	450	21	AAZ53171
41	68.8	4.7	8169	19	AAV26609
42	67.4	4.6	1119	24	ABQ90030
c 43	67	4.6	640681	24	ABA92787
c 44	66.8	4.6	450	21	AAZ53169
45	66.8	4.6	53789	19	AAV21187

ALIGNMENTS

RESULT 1  
AAZ47131  
ID AAZ47131 standard; DNA; 1450 BP.  
XX  
XX  
AC AAZ47131;  
XX  
DT 28-MAR-2000 (first entry)  
XX  
DE Pseudomonas aeruginosa murD gene.  
XX  
KW MurD; bacterial cell wall; biosynthesis; inhibitor; D-glutamate;  
KW UDP-N-acetylmuramyl-L-alanine precursor; ds.  
XX  
OS Pseudomonas aeruginosa.  
XX  
PN WO9961050-A1.  
XX  
PD 02-DEC-1999.  
XX  
PF 26-MAY-1999; 99WO-US11585.  
XX  
PR 29-MAY-1998; 98US-0087308.  
XX  
PA (MERI ) MERCK & CO INC.  
XX  
PI El-Sherbeini M, Azzolina B;  
XX  
DR WPI: 2000-072548/06.  
DR P-FSDB; AAY56047.  
XX  
PT New nucleic acid encoding the MurD protein of Pseudomonas aeruginosa,

N. meningitidis B  
Mycobacterium tube  
Mycobacterium tube  
Mycobacterium tube  
N. gonorrhoeae nuc  
DNA encoding novel  
DNA encoding novel  
DNA encoding novel  
DNA encoding novel  
DNA encoding novel  
DNA encoding novel  
DNA encoding novel  
Haemophilus influe  
Haemophilus influe  
Genomic fragment #  
DNA encoding novel  
DNA encoding novel  
Bacillus lichenifo  
C glutamicum codin  
Neisseria meningit  
Neisseria meningit  
Pseudomonas aerugi  
Nucleotide sequenc  
Pseudomonas aerugi  
S. cellulosum DNA  
Neisseria gonorrhe  
Platenolide syntha  
Platenolide syntha  
Neisseria meningit  
Actinomadura hibis  
M. capsulatus gene  
Buchnera sp. genom  
Neisseria gonorrhe  
Anycolatopsis medi



PR	23-MAY-2000; 2000US-206848P.	
PR	26-MAY-2000; 2000US-207727P.	
PR	23-OCT-2000; 2000US-242578P.	
PR	27-NOV-2000; 2000US-253625P.	
PR	22-DEC-2000; 2000US-257931P.	
PR	16-FEB-2001; 2001US-269308P.	
XX		
PA	(ELIT-) ELITRA PHARM INC.	
XX		
PI	Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GU;	
PI	Yamamoto RT, Xu HH;	
XX		
WPI:	2001-611495/70.	
P-PSDB:	AAU36416.	
DR		
XX		
PT	New polynucleotides for the identification and development of	
PT	antibiotics, comprise sequences of antisense nucleic acids -	
XX		
PS	Claim 27; Seq ID NO 7912; 51pp; English.	
XX		
CC	The invention relates to antisense inhibitors of genes essential to	
CC	prokaryotic cellular proliferation, their use in identifying the	
CC	genes, themselves in the discovery of novel antibiotics, the essential	
CC	genes themselves and the encoded proteins. The prokaryotes used are	
CC	Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella	
CC	pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The	
CC	invention is also useful for the identification of potential new targets	
CC	for antibiotic development. The antisense nucleic acids can also be used	
CC	to identify proteins used in proliferation, to express these proteins,	
CC	and to obtain antibodies capable of binding to the expressed proteins.	
CC	The proteins can be used to screen compounds in rational drug discovery	
CC	programmes. The antisense nucleic acid sequence is also useful to screen	
CC	for homologous nucleic acids which are required for cell proliferation in	
CC	a wide variety of organisms. The present sequence encodes an	
CC	essential prokaryotic cellular proliferation protein.	
CC	Note: The sequence data for this patent did not form part	
CC	of the printed specification, but was obtained in electronic	
CC	format directly from WIPO at	
CC	ftp.wipo.int/pub/published_pct_sequences.	
XX		
SQ	Sequence 1347 BP; 214 A; 453 C; 453 G; 227 T; 0 other:	

Query Match	92.9%	Score 1347;	DB 23;	Length 1347;
Best Local Similarity	100.0%;	Pred. No. 1.3e-233;		
Matches 1347;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0
QY	51	ATGAGCCTGATCGCCTCGGACCACTTCGCGATCGTTTCGCGCTCGCGAAGAGCGCGATG	110	
Db	1	ATGAGCCTGATCGCCTCGGACCACTTCGCGATCGTTTCGCGCTCGCGAAGAGCGCGATG	60	
QY	111	TCCCTGGTGGCGTAACTTGGCGCGCGCGGGCTTGCCCTTTCGCCCTGGTGCATACCCGAGAG	170	
Db	61	TCCCTGGTGGCGTAACTTGGCGCGCGCGGGCTTGCCCTTTCGCCCTGGTGCATACCCGAGAG	120	
QY	171	RACCGCGCGAGCTGGCCACCCCTCGGTGCCAGTATCCGCAGCTGGAAGTGGCTTCGGCG	230	
Db	121	RACCGCGCGAGCTGGCCACCCCTCGGTGCCAGTATCCGCAGCTGGAAGTGGCTTCGGCG	180	
QY	231	GAACTCGAGCGCGAGTTCCCTCTGCTCGCGCGCGGAACCTCTATGTACGCCCGGCTTGTGG	290	
Db	181	GAACTCGAGCGCGAGTTCCCTCTGCTCGCGCGCGGAACCTCTATGTACGCCCGGCTTGTGG	240	
QY	291	CTGGCGACCCCTGCGCTGGTACAGCGCGCGCGAAGGCGTGGCGATCTCCGGTGACATCT	350	
Db	241	CTGGCGACCCCTGCGCTGGTACAGCGCGCGCGAAGGCGTGGCGATCTCCGGTGACATCT	300	
QY	351	GATCTCTTCGCCCGCGAGCGGAAGGCCCGGATCTGCGCCATCACCGGTTCCACAGCGGAAG	410	
Db	301	GATCTCTTCGCCCGCGAGCGGAAGGCCCGGATCTGCGCCATCACCGGTTCCACAGCGGAAG	360	
QY	411	AGACCGGTGACCAACCCCTGGTGGGCGAATGAGCGGTGGCCGCGGACAAAGGTGTGCCCGGT	470	
Db	361	AGACCGGTGACCAACCCCTGGTGGGCGAATGAGCGGTGGCCGCGGACAAAGGTGTGCCCGGT	420	

Qy	471	GGCGGCAACCTTCGGCACCCCGGCGCTCGACCTGCTGCGCGACGACATCGAGCTGTAGCTG	530
Db	421	GGCGGCAACCTTCGGCACCCCGGCGCTCGACCTGCTGCGCGACGACATCGAGCTGTAGCTG	480
Qy	531	TTGGAGCTGTGCGAGCTTTCAGCTTGGAAACCTTCGCGATCGCTCAACGCCGAGGTGGCGACC	590
Db	481	TTGGAGCTGTGCGAGCTTTCAGCTTGGAAACCTTCGCGATCGCTCAACGCCGAGGTGGCGACC	540
Qy	591	GTGCTGAACGCTACGCGAAGACCATATGGATCGCTTACGACGGCATGGCTGACTACCACTG	650
Db	541	GTGCTGAACGCTACGCGAAGACCATATGGATCGCTTACGACGGCATGGCTGACTACCACTG	600
Qy	651	GCCAAGCACCGGATCTTCGGGGTGGCCGCCAGGTCTGTGTGATTCGGCCGATGCCCTG	710
Db	601	GCCAAGCACCGGATCTTCGGGGTGGCCGCCAGGTCTGTGTGATTCGGCCGATGCCCTG	660
Qy	711	ACCCGACCGGTGATCGCCGATACCGTGCCTGCTGCTTTCGGGCTTGAACAAGCCGGAC	770
Db	661	ACCCGACCGGTGATCGCCGATACCGTGCCTGCTGCTTTCGGGCTTGAACAAGCCGGAC	720
Qy	771	TTCGAAGCTTTTCGGCCTGTATCGAGGAAGACGGCCAGAGTGGCTTCCAGTTTCGAC	830
Db	721	TTCGAAGCTTTTCGGCCTGTATCGAGGAAGACGGCCAGAGTGGCTTCCAGTTTCGAC	780
Qy	831	AAGCTGCTGCCGTTGGCGAACTCAAAGTCCGTGGCGCCACACACTATTCCAAGCGCTC	890
Db	781	AAGCTGCTGCCGTTGGCGAACTCAAAGTCCGTGGCGCCACACACTATTCCAAGCGCTC	840
Qy	891	GCCGCGTGGCGTGGCCATGCGGCTCGGCTGCGCTTCGACGCCATGCTCGGCGCGCTG	950
Db	841	GCCGCGTGGCGTGGCCATGCGGCTCGGCTGCGCTTCGACGCCATGCTCGGCGCGCTG	900
Qy	951	AAGCGGTTTTCGGCCTGGCTCATPCGTGCCAGTGGGTACGCGAGCGCGAGGCGTGAGC	1010
Db	901	AAGCGGTTTTCGGCCTGGCTCATPCGTGCCAGTGGGTACGCGAGCGCGAGGCGTGAGC	960
Qy	1011	TACTACGACGATTCGAAGGCCACCAAGCTCGGCGCGCCCTTGGCGCGGATCGAGGGCTG	1070
Db	961	TACTACGACGATTCGAAGGCCACCAAGCTCGGCGCGCCCTTGGCGCGGATCGAGGGCTG	1020
Qy	1071	GGTGCCACATPCGACGGCAAGCTGGTGCTGCTCCGCGCGGAGACGCCAAGGGCGCCGAT	1130
Db	1021	GGTGCCGACATPCGACGGCAAGCTGGTGCTGCTCCGCGCGGAGACGCCAAGGGCGCCGAT	1080
Qy	1131	TTCCATGACCTTCGCGAGCCGGTCCGCGCTTTCGCCGGCGGTGGTACTGCTTGGCCGT	1190
Db	1081	TTCCATGACCTTCGCGAGCCGGTCCGCGCTTTCGCCGGCGGTGGTACTGCTTGGCCGT	1140
Qy	1191	GACGCCGGCTGATTTGCCAGGCACTGGGCAACCGGTACCGCTGGTGGCGGTGCGCAAGC	1250
Db	1141	GACGCCGGCTGATTTGCCAGGCACTGGGCAACCGGTACCGCTGGTGGCGGTGCGCAAGC	1200
Qy	1251	CTGACGAAGCAGTTCGCGAGGCCCGCGAGCTGCGCCGCGAAGCGGATCGGTGCTGTG	1310
Db	1201	CTGACGAAGCAGTTCGCGAGGCCCGCGAGCTGCGCCGCGAAGCGGATCGGTGCTGTG	1260
Qy	1311	TCGCCGGCTTCGCGAGCCTGGACATGTTCAAGAACTTCGAAGAACCGGACGCTTTC	1370
Db	1261	TCGCCGGCTTCGCGAGCCTGGACATGTTCAAGAACTTCGAAGAACCGGACGCTTTC	1320
Qy	1371	GCCAAAGCCGTAGAGGCTAGCGTGA	1397
Db	1321	GCCAAAGCCGTAGAGGCTAGCGTGA	1347

RESULT 3  
AAS56341  
ID AAS56341 standard; DNA; 1317 BP.  
XX  
AC AAS56341;  
XX  
DT 13-FEB-2002 (first entry)

XX Salmonella typhi DNA for cellular proliferation protein #374.  
 XX Antisense; ds; prokaryotic cellular proliferation gene;  
 KW antibiotic; antibacterial; drug design.  
 XX Salmonella typhi.  
 OS WO200170955-A2.  
 PN 27-SEP-2001.  
 PD 21-MAR-2001; 2001WO-US09180.  
 PF 21-MAR-2000; 2000US-191078P.  
 PR 23-MAY-2000; 2000US-206848P.  
 PR 26-MAY-2000; 2000US-207727P.  
 PR 23-OCT-2000; 2000US-242578P.  
 PR 27-NOV-2000; 2000US-253625P.  
 PR 22-DEC-2000; 2000US-257931P.  
 PR 16-FEB-2001; 2001US-269308P.  
 XX (ELIT-) ELITRA PHARM INC.  
 PA Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;  
 PI Yamamoto RT, Xu HH;  
 XX WPI: 2001-611495/70.  
 DR P-PSDB; AAU38482.  
 XX New polynucleotides for the identification and development of  
 PT antibiotics, comprise sequences of antisense nucleic acids -  
 XX Claim 27; Seq ID No 9978; 511pp; English.  
 XX The invention relates to antisense inhibitors of genes essential to  
 CC prokaryotic cellular proliferation, their use in identifying the  
 CC genes, their use in the discovery of novel antibiotics, the essential  
 CC genes, themselves and the encoded proteins. The prokaryotes used are  
 CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella  
 CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The  
 CC invention is also useful for the identification of potential new targets  
 CC for antibiotic development. The antisense nucleic acids can also be used  
 CC to identify proteins used in proliferation, to express these proteins,  
 CC and to obtain antibodies capable of binding to the expressed proteins.  
 CC The proteins can be used to screen compounds in rational drug discovery  
 CC programmes. The antisense nucleic acid sequence is also useful to screen  
 CC for homologous nucleic acids which are required for cell proliferation in  
 CC a wide variety of organisms. The present sequence encodes an  
 CC essential prokaryotic cellular proliferation protein.  
 CC Note: The sequence data for this patent did not form part  
 CC of the printed specification, but was obtained in electronic  
 CC format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences.  
 XX Sequence 1317 BP; 258 A; 341 C; 427 G; 291 T; 0 other;  
 SQ

Query Match 23.0%; Score 332.8; DB 23; Length 1317;  
 Best Local Similarity 56.1%; Pred. No. 3.6e-51;  
 Matches 747; Conservative 0; Mismatches 552; Indels 33; Gaps 5;

QY 69 GACCACTTCGGATCGTTCGGCGTGGCGAGAGCGGATGTCCTGGTGGCGTACCTG 128  
 DB 16 GACAAAAGGTGCTATTATCGTCTGGGCTTAACCGGACTCTCTTGGCTGGACTTTT 75  
 QY 129 GCGCGCGCGGTGCTTCGCGGTGTCGATACCGGAGAACCCCGGAGCTGGCC 188  
 DB 76 CTCGCCCGCGGTGACGCGCGGGGTGATGGATCTCGCGTGACGCGCGGTCTGGAT 135  
 QY 189 ACCCTGCGTGCCAGTATCCGAGGTGGAAAGTGGCGTTGGCGGAACTCGACGCCGAGTTC 248  
 DB 136 AAGTGGCG- - - - -CAAGAGTTGAGCGTCAAGTGGCGGCTGNAACGACGAGTGG 186

QY 249 CTCTGCTCCGCGCGGAACCTATATGTACGCCCGCGCTTGTGCTGCGACCCCTCGCGCTG 308  
 DB 187 CTCTTAGCGGGGATTTAATCTGCGCAGCCCTGGTATTCCTCTGCGCATCCTCTCTT 246  
 QY 309 GTACAGCGCGCGCGGAAGCGGTGCGCATCTCCGGTGACATCATCTCTGCGCCCGGAG 368  
 DB 247 AGCGCTGCGCGCGAGCGCTGGAGTGAATTTGTCGGCGATATCGAACTGTTTTCGCGGAA 306  
 QY 369 GCGAAGCGCGCATCTGCGCATCACCGTTCACACGCGAAGACCGGTGACACCCCTG 428  
 DB 307 GCGCAAGCGCGCATTTGTGCGCATCACCGCTTGACACGCGAAGACCGGTGACACCTTA 366  
 QY 429 GTGGGGAATGCGGTGCGCGACAAAGCGTGTGCGGTGCGCGCAACCTCTCGGCACC 488  
 DB 367 GTGGGCGAGATGCGGAAGCGCGGCGTCAATGTGCGGCTTGGCGCAATATCGCTCT 426  
 QY 489 CCGGCGCTCGACTGCTGCGCGACGACATCGAGCTGTAGTGTGTGGAGCTGTGCGAGCTTC 548  
 DB 427 CCGGCGCTGATGCTGTGATGCGCATCGGAATTTAGCTGTGTGAATTTATCCAGTTTC 486  
 QY 549 CAGCTGGAACCTGCGATCGCTCAACGCGGAGTGGCGACCGTGTGTAACGCTGACGCGAA 608  
 DB 487 CAATTGGAACCTACCTCAAGTTTTCAGCGCGCGCAACGGTGTCAACGCTCACTGAA 546  
 QY 609 GACCATATGATCGCTA- - -CGACGCGATGGCTGACTACCACTTGCGCCAAAGCACCAGATC 665  
 DB 547 GATCATATGACCGGTATCCGTTTGGTTTGAACAGTACCGGCGCGCAAACTGCGCGTC 606  
 QY 666 TTCCGCGGTGCGCGCGAGTGTGGTGAATCGCGCGATGCGCTGACCGCGCGCTGATC 725  
 DB 607 TAGGAGAGGCGAAGTGTGCTAGTGAATGCGGATGACGCGTGTGACGATGCGCGTACGC 666  
 QY 726 GCGCATACCTGCGCTGCTGCTGGCGCTGAACAAGCGCGACTTCAAGGCTTTCGCG 785  
 DB 667 GGGCGCGATGAGCGCTGCGTCACTTGGCGTCAATATGGTGTATATCA- - - - -C 717  
 QY 786 CTGATCGAGAAAGCGCGAGAGTGGTGGCGTTCAGTTCGACAAGTGTGTCGGGT 845  
 DB 718 CTTAATCTGTCAGCAGGCGGAAACCTGGCTACGAGTCAAAAGGTGAGAAGGTGTGAACGTG 777  
 QY 846 GCGCAACTCAAGATCGTGGCGCGCCCAACTTCCAAACGCGCTCGCGCGTGGCGCTG 905  
 DB 778 AAGAGATGAAGCTTCCCGTTCAGTAACTATACCAATGCGTTACGCGCGCTGGCGCTG 837  
 QY 906 GGCCATGCGGTGCGCTGCGCTGTCGACGCGCATGCTGCGCGCGCTGAAGCGTTCGCGC 965  
 DB 838 GCGGATGCGGTAGTGTGCGCGCGCGCAGCAGTTTGAAGCGCTGTGACGACATTTACCGGT 897  
 QY 966 CTGGCTCATCGTTCGAGTGGGTACCGAGCGCGGCGGTGAGCTACTACGACGATTC 1025  
 DB 898 CTGGCGCATCGCTTCAGCTGGCGTTGAGCATAAACGGGCTTCGTTGGATCAACGACTCG 957  
 QY 1026 AAGGCCACCAAGCTGCGCGCGCTTGGCGCGATCGAGGGCTGGTCCGACATCGAC 1085  
 DB 958 AAGCGCAACATGTCGCGAGTACCGAAGCGCGCTTAACGGTTTGCATG- - - - -TGGAC 1011  
 QY 1086 GGCAAGCTGCTGCTCGCGCGGAGAGCGCAAGCGCGCGATTTCCATGACCTGCGC 1145  
 DB 1012 GGTACGCTCATCTGCTGCTGCGCGCGCAGCGTAACTGCGCAGACTTTTCTCCGCTGGCG 1071  
 QY 1146 GAGCGGTGCGCGCTTCTGCGCGGGGTGCTGCTGCTGGCGGTGACCGCGCGGTGAT 1205  
 DB 1072 CGCTATCTGACCGGGATCGTATCGCGCTGTATTGCTTTGGCGGCGATGGCGCGACT 1131  
 QY 1206 GCGCAGGCACTGGCAACGCGGTACCGTGTGCGGTGCGCAACCTGACCAAGCAGT 1265  
 DB 1132 GCGGCACTGC- - - - -GTCCGAAATGCGCCCAACAGACTGACGATGGAAGGCGATG 1185  
 QY 1266 GCGCAGGCGCGGAGCTGGCGCGGAAAGCGATGGGTGCTGTTGTGCGCGCGCTGCGCG 1325  
 DB 1186 CGTTTGTGCGCGCGGCTTACGCGCGGTGATGTTGCTGTTGTGCGCGCGCTGCGCG 1245  
 QY 1326 AGCCTGGACATGTTCAAGAACTTCGAAGAACCGCGGACGCTGTTCCGCCAAGACCGTAGAG 1385

[illegible]









```
Db 52199 GCTTTTGGAGGCGAATGCGACGCGAAGGCAAAAAGCGGACGTGTGGTGTTCGAGCT 52140
QY 539 GTCGAGCTTCCAGCTGGAACCTCGATCGCTCAACGCGGAGGTGGCGACCGTCTGAA 598
Db 52139 TTCCAGCTTCCAACTGGAACACACCGAAGCGCTCGCTCGAGCTGCGGCGGAGGTGCTGAA 52080
QY 599 CGTCAGCGAAGACCATATGATGCTAGCAGCGCATGCTGACTACACCTGGCCAAAGCA 658
Db 52079 CATTTCGAGAGACCATCGACCGCTACGACGATGCTGCACTATGCGCATACCAAGC 52020
QY 659 CCGGATCTTCCCGGTGCCCGCAGGTGCTGTGTGAATCGCGCCGATGCCCTTGACCGACC 718
Db 52019 CAAGATTTTCCGTGCGCAGCGGTGCTGAGTTTGAATGCGGACGATGCGTTCGCGCGC 51960
QY 719 GCTG-ATGCGCATACCGTGGCGTGTGCTGCTCGGCTGAACAAGCGGACTCAAGG 777
Db 51959 GATGAAGCGTCCGCGCGGAGGTAAATGTTTTGCTTGGAAACAGAGCTGATTTCTG 51900
QY 778 CTTTCGGCTGATCGAGGAAGCGCCAGAGTGGCTGGCTTCCAGTTTCGACAAAGCTGC 837
Db 51899 GTTGAACCGAGACAGCGCGCTGAACAAGGCAATGAAG-----ATTGGA-51853
QY 838 TCCCGGTTGGCACTAGATCCGTGGCGGCCCACTATTTCCAAAGCGCTCGCGCGC 897
Db 51852 TTGTACCGAAGACATTCGTTGCAAGGTCTGCACACCGCGCTAACGTCATGCTGGTGGC 51793
QY 898 TGGCGCTGGGCCATCGGTGGCGCTGCGCTTCCAGCCATGCTCGCGCGCTGAAGGCGT 957
Db 51792 TGGCTTTGTGAGGCCATCGTTTGTGCGCGAAGCATTGTCTCAACACGTCAAACT 51733
QY 958 TTTCCGGCTGCTCATGCTGCCAGTGGTACGCGAGCGGCGGCTGAGCTACTACG 1017
Db 51732 TCCAAAGGCTCGCGCACCGCTGGGAAATAATCGCGGAGAAAACGCGTGGTGTATTATCG 51673
QY 1018 AGATTTCAAGCCACCACTCGCGCGCGCTCGCGCGCTGCGCGGATCGAGGGCTGGTGGC 1077
Db 51672 ACGACAGCAAGGCGACGATGTCGCGCGGACTGCGCGCGGAT-----TGCGG 51625
QY 1078 ACATCGACGCGAAGCTGCTGCTGCTCGCGCGGAGACGCGAAGGCGCGGATTTCCATG 1137
Db 51624 GTTTCGAAATCCGCTCTTCGTGATTTTGGCGGCGATGGTAAAGGCGAGGACTTCACG 51565
QY 1138 ACTCGCGAGCGGTGCGCGCTTCTGCGCGCGGTGCTGCTGCGCGGTGCTGCGCGGTGCGG 1197
Db 51564 CCGTCGCGATCACTGTGAGCAAGGCAAAAAGCGGTGTTCTTGTATGCTGCTGATGCGC 51505
QY 1198 GCTGATTCGCCAGGCACTG---GGCAAGCGGTACCGCTGCTGCGCGCTCGCAAGCTGG 1254
Db 51504 CGCAATCCGCGCGATTTGGACGCTGCGGCTTGAATATGACCGACTGCGCACTTTGG 51445
QY 1255 ACGAAGCACTCGCGCAGCGCGCGAGCTGGCCCGGAGCGGATGCGGTGCTGTTGTGTCG 1314
Db 51444 GAGAAGCGCTTCAGACGGCATATGCCAAGCGGAGCGGCGATATGTTGTTGCTCAGCG 51385
QY 1315 CGCGCTCGCGAGCTGACATGTTCAAGAACTTCAAGAAACGCGGACCGCTGTTGCGCA 1374
Db 51384 CGCGCTCGCGAGCTTGTATGTTCAAGAGGCTACGCGCACCGTTCGGAAGTGTATTATCG 51325
QY 1375 AAGCGGTAGAGG 1386
Db 51324 AAGCGTTAAGG 51313
```

## RESULT 8

AAF21607

ID AAF21607 standard; DNA; 349980 BP.

XX AAF21607;

XX 13-MAR-2001 (first entry)

DT

XX Neisseria meningitidis B nucleotide sequence SEQ ID NO:108.

DE

```
XX Neisseria meningitidis; Neisseria gonorrhoeae; immunogenic; vaccine;
KW diagnosis; antigen; detection; infection; gene therapy; antibacterial;
KW ds.
XX Neisseria meningitidis.
XX WO2000066791-A1.
XX 09-NOV-2000.
XX 08-MAR-2000; 2000WO-US05928.
XX 30-APR-1999; 99US-0132068.
XX 08-OCT-1999; 99WO-US23573.
XX 28-FEB-2000; 2000GB-0004695.
XX (CHIR ) CHIRON CORP.
XX (GENO-) INST GENOMIC RES.
XX Pizza M, Hickey E, Peterson J, Tettelin H, Venter JC, Maignani V;
PI Galeotti C, Mora M, Ratti G, Scarselli M, Scarlato V, Rappuoli R;
PI Frazer CM, Grandi G;
XX WPI; 2000-647603/62.
XX Neisseria meningitidis B full length genome sequence and open reading
frames are used to detect, treat and prevent Neisserial infections .
XX Claim 7: Appendix A; 692pp; English.
XX The present invention describes the full length genome of
Neisseria meningitidis B (NMB). The sequences in AAF21544 and AAF21607
to AAF21613 represent fragments of the NMB genomic sequence, as the
sequence was too long to go in a record on its own it was split into 8
sequences which overlap each other at the beginning and end of each
sequence by 49980 bp (i.e. the last 49980 bp of AAF21544 is repeated at
the beginning of AAF21607, the last 49980 bp of AAF21607 are repeated at
the beginning of AAF21608, and so on). AAF21545 to AAF21588 encode the
Neisseria proteins given in AAB58550 to AAB58593, and AAF21589 to
AAF21606 represent PCR primers which are used in the exemplification of
the present invention. The NMB genome and fragments from it have
antibacterial activity, and can be used in vaccines and gene therapy.
Neisseria nucleic acids, proteins and/or antibodies which binds to the
proteins can be used in compositions for treating or preventing infection
due to Neisserial bacteria or as a diagnostic reagent for detecting the
presence of Neisserial bacteria or of antibodies raised to Neisserial
bacteria. Computers, computer memory, computer storage medium or computer
databases can be used in a search to identify open reading frames (ORFs)
or coding sequences within the NMB genome. The DNA sequences provide
further opportunities to find antigenic or immunogenic proteins which are
more effective in vaccines than the outer membrane proteins currently
used.
```

Sequence 349980 BP; 84410 A; 84863 C; 94187 G; 86520 T; 0 other;

Query Match 12.4%; Score 180; DB 21; Length 349980;  
Best Local Similarity 49.7%; Pred. No. 1.1e-23;  
Matches 662; Conservative 0; Mismatches 620; Indels 50; Gaps 6;

```
QY 80 CATCGTTTCGCGCTCGGCAAGAGCGGATGCCCTGTGTGCGGTACCTGCGCGCGCGCGG 139
Db 128568 CCTCGTCGCGGACTCGCGGTACGGGTATTTCCATGATTTGCTCGCAAAAACGG 128627
QY 140 CTTGCTTTTCGCGGTGCTGATACCGAGAGAACCGCGGAGCTGCCACCTCGGTGC 199
Db 128628 CCGGAGGTGCTGCTGATGCTGCGGAGCTGAAGCGGAGACCGGTTCGCAATCGGTAA 128687
QY 200 CCAGTATCCGAGGTGGAAGTTCGCTGCGGAGCTGAAGCGGAGACCGGTTCGCTGCTCGCG 259
Db 128688 GATGTTGACGGTGTGCTGTTTACACGCGCGCTCTGAAGATGCGCTGGACACCGGTTT 128747
QY 260 CCGCGAACTCTATGTAGCCCGCGCTTGTGCTGCGCACCCCTGCGCTGGTACAGCCCGC 319
```

128748	CGATATTCCTGGCTCTCATGTCCTCCGGCATCAGCGAGCGGACGCCGATATTTGAGGCGTTCAA	128807
320	CGCGAAGGCGTGGCGATCTCCGGTGACATCGATCTCTTGGCCGCGAGGCGCAAGGCCGCC	379
128808	GCAAAACGGCGAGCGGTGTGGGGGACATCGAATTGCTGGCGACATGTGTGAACCGCGG	128867
380	GATCGTC-----GCCATCACCGGTTCCAAACGGAAGAGCACCGGTGACCAACCTGGT	430
128868	GGACGACAAAGTAATTGCGATTATTCGGGACGACGCAAAACACCGGTAAACGAGCTGT	128927
431	GGGCGAATGCGGTTGGCCGGGACAAACGCTGTCGCCGTGCGGCGCAACCTCGGCACGCC	490
128928	CGGCTATCTCTGATCAAGTCGGGCTCGATACCGTTATCGGGCGCAATATCGGCACGCC	128987
491	GGCGCTCG-----ACCTGCTGCCCGACACATCGAGCTGTACGTTGTTGGAGCT	538
128988	GGTTTTGGAGCGGAATGGCAGCGCGAAGGCAAAAGCGGACGTGTGGGTGTGGAGCT	129047
539	GTCGAGCTCCAGCTGGAAACCTCGGATCGCTCAACGCCGAGGTGGCGACCGTCTGTA	598
129048	TTCAGCTTCCACTGGAAACACCGAAAGCGTGGTCCGACTGGCGGAGCGGTGCTGAA	129107
599	CGTCAGCGAAGACCATATGGATCGCTACGACGGCATGCTGACTACCACTGGCCAAAGCA	658
129108	CATTTCCGAAGACCATCTCGACCGCTACGACGACTTGTCTGACTATGCGCATACCAAGC	129167
659	CCGATCTTCGGCGGTGCCCGCAGGTGCTGGTGAATCGCGCGCATGCCCTGACCCAGCC	718
129168	CAAGATTTTCGGTGGCGACGCGTGCAGGTTTGAATCGGACGATGCGTTCTGCGCGCG	129227
719	GCTG-ATTCGCCGATACCGTGCCTGCTGGTTCGGCTGAAACGACCGGACTTCAAGG	777
129228	GATCAAGCGTCCGCGCGCAGGTAAATGCTTTTCGTTGGAACACGAAGCTGATTTCTG	129287
778	CTTTCCGCTGATCGAAGACAGCGCCAGAGTGGCTGGCGTTCCAGTTTCGACAAAGTGC	837
129288	GTTGGAACGGGAGACAGCGCCCTGAAACAGGCAATGAAG-----ATTTGA	129334
838	TGCCGGTTGGGAACTGAAGATCCGTGGCGGCCACAACTATTCAAACGCGTCGCGCGCG	897
129335	TTGTCACGCAAGACATTCGGTTGCAAGGTCTGCACACGCGCGCTAACGTCATGGCTGCC	129394
898	TGGCGCTGGGCGATCGGTCGGCTCGCGTTTCGACGCGCATGCTCGCGCGCTGAAGCGGT	957
129395	TGGCTTTGTGTAGGCGCATCGGTTGTGCGCGAAGCATTTCTCGAACACGCTCAAAACCT	129454
958	TTTCCGCGCTGGCTCATCGCTGCCAGTGGTACGCGAGCGGCGGCGTGAGCTACTACG	1017
129455	TCCAAAGGCTCCCGACCGGTGGAAAAATCGCGGAGAAAAATCGCGGAGAAAAATCGG	129514
1018	ACGATTCGAAGGCCAACACGTCGCGCGCGCCCTTGGCGGCGATCGAAGGCGTGGGTGCCG	1077
129515	ACGACAGAAAGGACGAAATGTCGCGCGGACTGCGCGCGCAT-----TGCCG	129562
1078	ACATCGAGCGCAAGCTGTCGCTCGCGCGGAGAGCGCAAGGCGCGCATTTCCATG	1137
129563	GTTTGCAAAATCCGCTCTTCGTGATTTGGCGGCGATGGGTAAAGGCGAGGACTTCACGC	129622
1138	ACCTGCGGAGCCGGTCGCGCGCTTCTCGCGGCGGTGGTACTGCTTGGCGGTGACGCGG	1197
129623	CCCTGCGCATGCACTGGTAGGCAAGCAAAAGCGGTGTTCTTCATTGGTTCGATCGCG	129682
1198	GGCTGATTTCCAGGCACTG---GGCAACGGGTACCGTGTGCGGCTGCGCAACGCTGG	1254
129683	CGAAATCCCGCCGCGATTTGGACGGCTGCGGCTTGAATATGACCGAGTGGCCCACTTGG	129742
1255	ACGAAGCAGTCCCGGACGCGCGAGCTGGCGCCGCAAGGCGATGCGGTGCTGTGTCG	1314
129743	GAGAACCGTTACAGACGGCATATGCCCAAGCCGAGCAGGCGATATTGTTGTTGCTCAGCC	129802
1315	CGGCTTCGCGAGCCTGGACATGTTCAAGAACTTTCGAAGAACCGCGGACGCTGTTTCGCCA	1374

Db	129803	CGCCTGCGCGAGCTTTGATATGTTCAAAAGGCTACGCGACCGTTTCGGAAGTGTATTATCG	129866
QY	1375	RAGCGGTAGAGG	1386
Db	129863	RAGCGTTTAAAGG	129874
RESULT 9			
AAA81490			
ID	AAA81490	standard; DNA; 1437668 BP.	
XX	AC	AAA81490;	
XX	04-DEC-2000	(first entry)	
DE	N. meningitidis B	full length genome DNA sequence SEQ ID NO:1068.	
KW	Neisseria meningitidis; Neisseria gonorrhoeae; genome; immunogenic;		
KW	antigen; vaccine; diagnosis; infection; antibacterial; identification;		
KW	Meningococcus B; MenB; ds.		
XX	OS	Neisseria meningitidis.	
XX	WO200022430-A2.		
PN	20-APR-2000.		
XX	08-OCT-1999;	99WO-US33573.	
PF	09-OCT-1998;	98US-0103794.	
XX	30-APR-1999;	99US-0132068.	
PR	(CHIR )	CHIRON CORP.	
XX	Frazer CM, Hickey E, Peterson J, Tettelin H, Venter JC;		
PI	Maignani V, Galeotti C, Mora M, Ratti G, Scarselli M, Scarlato V;		
PI	Rappuoli R, Piza M;		
XX	WPI; 2000-318079/27.		
DR	Isolated nucleotide sequences of Neisseria meningitidis which can be		
XX	used in the diagnosis and treatment of N. meningitidis infection and		
PT	other Neisserial infections, for example, N.gonorrhoea -		
PT	Claim 7; Page 866-1272; 1760pp: English.		
XX	The present invention describes methods of obtaining immunogenic		
CC	proteins from Neisseria genomic sequences. AAA81453 to AAA82414		
CC	represent specifically claimed Neisseria meningitidis genomic DNA		
CC	sequences; AAA81260 to AAA81303 and AAB25620 to AAB25663 represent		
CC	Neisseria DNA sequences and their corresponding proteins; AAA81254 to		
CC	AAA81259 and AAA81304 to AAA81321 represent PCR primers used in the		
CC	isolation of Neisseria meningitidis DNA sequences; and AAA81322 to		
CC	AAA81452 represent Neisseria meningitidis MenB polynucleotide ORF		
CC	sequences, which are all used in the exemplification of the present		
CC	invention. The nucleic acid sequences, protein sequences, and antibodies		
CC	against them, can be used in the manufacture of a composition. The		
CC	composition can be used as a medicament (or in the manufacture of a		
CC	medicament) for treating, preventing or diagnosing infection due to		
CC	Neisserial bacteria. For example, some of the identified proteins could		
CC	be components of vaccines against Meningococcus B; against all serotypes;		
CC	and/or against all pathogenic Neisseriae. Identification of sequences		
CC	from the bacterium will also facilitate production of biological probes,		
CC	particularly organism-specific probes. Attempts to make efficacious		
CC	Meningococcus B vaccines have failed mainly due to antigen tolerance.		
CC	Multivalent vaccines have also been tried but none have successfully		
CC	overcome antigenic variability. The provision of further, complete		
CC	sequences may provide an opportunity to identify secreted or surface		
CC	exposed proteins that may be presumed targets for the immune system and		
CC	which are not antigenically variable or at least more conserved than		
CC	other more variable regions.		
XX	Sequence 1437668 BP:344338 A: 353206 C: 385074 G: 355045 T:5 other;		
SQ			

Query Match	12.4%	Score 180;	DB 21;	Length 1437668;
Best Local Similarity	49.7%;	Prod. No. 1.1e-23;		
Matches 662;	Conservative	0;	Mismatches 620;	Indels 50; Gaps 6;
Qy	80	CATCGTTGTTCGGCTTCGGCAAGACGGCATGTCCTGTGTGGCTACCTTGGCGCGCGCGG	139	
Db	428568	CCTCGTCGGCACTCGGGGTACGGGTATTTTCCATGATGCCYACCTGCGCAAAACGG	428627	
Qy	140	CTTGCCCTTTTCGGCTGTCGATACCCGAGAACCCCGCGGAGCTGGCCACACCTCGGTGC	199	
Db	428628	CGCGAGGTTCTCGTATGATCGGAGCTGAAGCCGGAACGCGTGTGCAAAATCGGTAA	428687	
Qy	200	CCAGTATCCGCAAGTGAAGTGCCTTGGCGGAACTCGACGCCGAGTTCTCTGCTCCGC	259	
Db	428688	GATGTTGACGGGTTGGTGTTTACACGGCGCTCTGAAGATGCGCTGGACAACGCTTT	428747	
Qy	260	CCGCGAACTCTATGTCAGCCCCGGCTTGTGCTGCGCACCCCTGCGCTGGTACAGGCCGC	319	
Db	428748	CGATATTCTTGCTCTCAGTCCCGCATCAGCGAGCGGACGCGATATTAGGCGTTCAA	428807	
Qy	320	CGCGAAAGCGTGCATCTCCGGTGACATCGATCTCTTCGCCCGGAGCGAAGGCCCC	379	
Db	428808	GCAAAACGGCGGACCGGTGTGGCGGACATCGAATGCTGGCGGACATGTGAACCGCGG	428867	
Qy	380	GATCGTC-----GCCATCACCGGTTCCAACGCGAAGACACCGTACACACCTGTGT	430	
Db	428868	GGACGACAAGTAATTGCGATTACGGCGACGAACGCAAAACACCGGTAAACGAGCTGT	428927	
Qy	431	GGGCGAAATGGCGGTGGCCGGGACACGCTGTCCCGTGGCGGCGAACCTCGGCACCCC	490	
Db	428928	CGGTATCTCTGTATCAAGTCCGGCTGGATACCTTTATCGCGGCAATATCGGACGCG	428987	
Qy	491	GGCGCTCG-----ACCTGCTGGCGACGACATCGAGCTGACGTGTTGAGCT	538	
Db	428988	GGTTTGGAGCGGAAATGGACGCGGCAAGGCAAAAGGCGACGTGTGGGTGTTGAGCT	429047	
Qy	539	GTCAGCTTCCAGTGGAAACCTGCGATCGCTCAACGCCGAGGTGGCGACCGTGTGAA	598	
Db	429048	TTCCAGCTTCCAACGTGGAACACCGGAAAGCTGCGTCCGACTGCGGCGACGCTGTGA	429107	
Qy	599	GTGACGCAAGACCATATGGATCCGTACGAGGCGATGCTGACTACCACCTGGCCAAAGCA	658	
Db	429108	CATTCCGAAGACCATCTCGACCGCTACGAGACCTTGTCTGCAATGCGCATACCAAAGC	429167	
Qy	659	CCGATCTTCGCGGTGCCGCCAGGTGCTGGTGAATCGCGCGATGCGCTGACCCGACG	718	
Db	429168	CAAGATTTCCGTGCGACGCGTGCAGTTTGAATCGGACGATGCTTCTGCGCGC	429227	
Qy	719	GCTG-ATCGCCGATACCGTGCCTGCTGCTGTTCCGGCTGAAACGCGGACTTCAAGG	777	
Db	429228	GATGAAGCGTCCGCGCGAGGTAAATGTTTTCTGTTGGAACACGAAGCTGATTTCTG	429287	
Qy	778	CTTTCGGCTGATCGAGGAACAGCGCCAAAGTGGCTGGCGTTCCAGTTCGCACAAGCTCG	837	
Db	429288	GTTGGAACGAGACAGGCCCTCTGAACAAGGCAATGAAG-----ATTGA	429334	
Qy	838	TGCGGTTGGCGAACTGAAGATCCGTGGCGGCCACAACATATTCACACGCGTCCGCGCG	897	
Db	429335	TTGTCAAGAGACATTCGGTTGCAAGGCTGCAACAAGCGCTAACGTCATGGTGC	429394	
Qy	898	TGGCGCTGGCCATCGGTGCGCTGCGTTTCGAGCCCATGCTCGCGCGCTGAAGCGT	957	
Db	429395	TGGCTTGTGTGAGGCCATCCGTTGTTCGCGGAAGCATGTCTCGAACACGTCFAAACCT	429454	
Qy	958	TTTCCGGCTTGCTCATCGCTGCCAGTGGGTACGCGAGCGGAGGCGCTGAGCTACTAGC	1017	
Db	429455	TCCAAGGCTCCGACCGCTGGNAANAATCGGCGAANAACGCGTGTGTTATCG	429514	
Qy	1018	ACGATTCGAAGGCCAACACTGTCGCGCGCCGCTGGCGGCGATCGAGGGCTGGGTGCGG	1077	
Db	429515	ACGACAGAAAGGACGAATGTGCGGCGACTGCGCGCGCAT-----TGCGG	429562	

Qy	1078	ACATCGAGGCAACTGGTGTCTGC	CCGCCGGAGACGGC	CAAGGGCCCGCATTTCCATG	1137	
Db	429563	GTTTGCAAAATCCGCTCTTCGTG	ATTTTGGCGGCATGGGTAAAGGG	CAGGACTTCA	429562	
Qy	1138	ACCTGCGGAGCGGTGCGCGGCTTC	TCCCGGCGGTGTACTCTCTTGGCCG	TGACGCGG	1197	
Db	429623	CCCTGCGCGATGCACTGGTAGGCA	AGGCAAAAAGGCGTGTCTTATTGGT	GTCTGATGCGC	429682	
Qy	1198	GGCTGATTGCCAGGCAC	TG---GGCAACGCGGTACC	GCTGGTGC	CGCGCTCGCAACGCTGG	1254
Db	429683	CGCAAAATCCGCGGATTTGGAC	GGCTGCGGCTTGAATATCACCGA	CTGCGCCACTTTGG	429742	
Qy	1255	ACGAAGCAGTCGCGCAGCCCGC	CGAGCTGGCCCGGAAGGCGATCG	GGTGTGTGTGTCG	1314	
Db	429743	GAGAAGCGGTTCAGACGGCATAT	GCCCAGCGCAAGCAGCGGATATT	GTTGTGCTCA	429802	
Qy	1315	CGGCTGCGGAGCGCTGACCATGT	TCAAGAACAATTCGAAGAACG	CGGACGCTGTTCG	1374	
Db	429803	CCGCTGCGGAGCTTGTGATGAT	GTTCAAAGGCTACGCGCACCG	CTTCGGAAGTGTTATCG	429862	
Qy	1375	AAGCCGTAGAGG	1386			
Db	429863	AAGCGTTAAGG	429874			
 RESULT 10 AAH52025 ID AAH52025 standard; DNA; 1461 BP.						
Xx	AAH52025;					
Ac	04-SEP-2001 (first entry)					
Dt	Mycobacterium tuberculosis potential drug target gene SEQ ID 79.					
De	Drug target; growth; organism viability; characterisation; ds.					
Kw	Mycobacterium tuberculosis.					
Os	WO200135317-A1.					
Pn	17-MAY-2001.					
Pd	13-NOV-2000; 2000WO-US31152.					
Pf	12-NOV-1999; 99US-0165086.					
Pp	12-NOV-1999; 99US-0165124.					
Pr	01-FEB-2000; 2000US-0179531.					
Px	(REGC ) UNIV CALIFORNIA.					
Pa	Eisenberg D, Rotstein SH, Marcotte EM;					
Pi	WPI; 2001-329193/34.					
Pj	P-PSDB; AAG81174.					
Pk	Identifying nucleotide or polypeptide sequence for use as drug target,					
Pt	involves providing algorithm that analyzes a functional relationship					
Pt	between nucleotide or polypeptide sequences, and comparing the					
Pt	sequences -					
Px	Disclosure; Page 100; 207pp; English.					
Px	This invention relates to a method for identifying a nucleotide or					
Px	polypeptide sequence that may be a drug target, or essential for growth					
Px	or viability of an organism. Polynucleotide sequences AAH51947 - AAH52092					
Px	represent DNA encoding proteins AAG81096 - AAG81241, Mycobacterium					
Px	tuberculosis proteins which are potential drug targets. The DNA and					
Px	protein sequences are used to illustrate the method of the invention. The					
Px	method involves providing an unknown nucleotide or polypeptide sequences,					
Px	and comparing it to a number of sequences along with at least one					
Px	algorithm capable of analysing a functional relationship between					
Px	nucleotide and polypeptide sequences. The method is useful for					
Px	identifying a nucleotide or polypeptide sequence for use as drug target,					
Px	involves providing algorithm that analyzes a functional relationship					
Px	between nucleotide or polypeptide sequences, and comparing the					
Px	sequences -					
Px	Disclosure; Page 100; 207pp; English.					

DB 1099 GCGATGGCGTCGCGGCTGGTCGGTGCCTGATCGGCCGGGATCGCGCAGCGGTGCC 1158

2110 45

[illegible]

QY 432 GCGAAATGCGGTGCGCGGCAAGAGTGTGCGCGGCAACCTCGGCACCCG 491  
 DB 2415999 CAGCCATGCTGATCGCGCGGTGCGCGCGCGGTGCTGCGGCAATATCGCAGTGG 2415940  
 QY 492 GCGCTGACCTGCTGCGCGGACGACATCGAGCTGTGAGTGTGAGCTGTCCAG 551  
 DB 2415939 GTGCTGGATGTGCTGGAGGAGCGCGGAGTGTGCGCGGTGAGTGTCCAGTTCCAG 2415880  
 QY 552 CTGGAAACCTGCGATCGCTCAACCGCGAGGTGGCGCGGTGCTCAAGCTCAGCGAAGC 611  
 DB 2415879 CTGCACTGGCGCCCTGCTGCGCGCGCGGCGCGGTGCTCAACATTCGCCAAGAC 2415820  
 QY 612 CATATGATGCTGCTAGCGGCTGCTGACTACCACTGGCGCAACCGCGATCTTCGCG 671  
 DB 2415819 CACCTGGAGTGCATGCGCACGATGCGCAATACCGCGGCAAGCGCGGTGTGACC 2415760  
 QY 672 GGTGCGCGGAGGTGCTGCTGAGTGAATCGCGCGGATGCGCGGAGTGTGATCGCGGAT 731  
 DB 2415759 GCGCGGGTAGCGGTGCGCGGCTGATGACAGCGGCGCGGCTGCTGCGACGCTCA 2415700  
 QY 732 ACCGTGCGGTGCTGCTGCG---TTGCGCTGACACAGCGGACTTCAAGCTTTCGCGCTG 788  
 DB 2415699 CCGGCGAGGTGCGGTGCGGTTCGCGTTCGCGAGCGCGCGCGGGAATGCGCGTG 2415640  
 QY 789 ATCGAGGAAGCGCGGAGGAGTGGCTGCGGTTCCTGAGTTCGACAAGCTGCTGCGGTGGC 848  
 DB 2415639 CCGGAGCGCCACTGGTGCATGCGCGCTTCCTCGAGACTGACGCTGCTGCGGTGCG 2415580  
 QY 849 GAACCTGAAGATCCGTGGCGGCGCACAACTATTCCACAGCGCTGCGCGCGTGGCGTGGC 908  
 DB 2415579 TCGATACCGGTGCCAGGTCCGCTGCGGTGCTGAGCGCTGCGCGCGCGGTGGCC 2415520  
 QY 909 CATGCGGTGCGCTGCGGTTCGACCGCATGCTGCGCGCGCTGAAGGCGTTTTCGCGCTG 968  
 DB 2415519 CGCTCGGTGCGGTGCGCGCGGTGCGATGCGCGGCGCGGTACGCTGTTGAGTGGCG 2415460  
 QY 969 GCTCATGCTGCTGCTGCGCGGAGAGCGAGCGGCGGTGAGTGTGCTGCTGAGTTCATGAC 1028  
 DB 2415459 CGACACCGCGCGAGGTGTGCGCGCTTCCGAGCGCATCACTAGTGGAGACTCCAG 2415400  
 QY 1029 GCCACACAGTGGCGCGCGCTGCGCGGATGCGAGGCGGTGGTGGCGGATGCGAGCGCGC 1088  
 DB 2415399 GCCACCAACCGCGCGCGCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2415355  
 QY 1089 AAGCTGTGCTGCTGCGCGGAGAGCGAGCGGCGGTGCTGCTGCTGCTGCTGCTGCTGCTG 1148  
 DB 2415354 AGGTGTGATGATGCGCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2415295  
 QY 1149 CCGGTGCGCGCTTCTGCGCGCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1208  
 DB 2415294 GCGATGGCGTGGCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2415235  
 QY 1209 GAGGC 1213  
 DB 2415234 GAGGC 2415230  
 RESULT 12  
 ID AA199683/c standard; DNA; 4403765 BP.  
 AC AA199683;  
 DT 15-JAN-2002 (first entry)  
 XX Mycobacterium tuberculosis strain H37Rv genome SEQ ID NO 2.  
 DE Mycobacterium tuberculosis; strain H37Rv; patient treatment; epidemic monitoring; ds.  
 KW Mycobacterium tuberculosis; strain H37Rv; patient treatment; epidemic monitoring; ds.  
 OS Mycobacterium tuberculosis.

US294328-B1.  
 25-SEP-2001.  
 24-JUN-1998; 98US-0103840.  
 24-JUN-1998; 98US-0103840.  
 (GENO-) INST GENOMIC RES.  
 Fleischmann RD, White OR, Fraser CM, Venter JC;  
 WPI; 2001-647361/74.  
 Evaluating strain variation of Mycobacterium tuberculosis, comprises determining the nucleotide sequence of the strain at positions in the genome corresponding to positions where M. tuberculosis strains CDC 1551 and H37Rv differ  
 Claim 4; SEQ ID NO 2; 3pp + Sequence Listing; English.  
 The invention relates to evaluating strain variation within and between different populations of the tuberculosis bacterial pathogen, Mycobacterium tuberculosis or related Mycobacterium by determining the nucleotide sequence of the first strain at positions in the complete sequence of the genome that correspond to positions that differ in the nucleotide sequences of M. tuberculosis strains CDC 1551 (AA199683) and H37Rv (AA199682). The method is useful for evaluating strain variation of M. tuberculosis and has valuable application in the fields of tuberculosis genetics, epidemiology, patient treatment and epidemic monitoring.  
 Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from USPTO at seqdata.uspto.gov/sequence.html?DocID=6294328B1.  
 Sequence 4403765 BP; 757105 A; 1447799 C; 1441301 G; 757371 T; 189 other;  
 Query Match 11.9%; Score 172.2; DB 22; Length 4403765;  
 Best Local Similarity 51.6%; Pred No 2.7e-22;  
 Matches 498; Conservative 0; Mismatches 428; Indels 39; Gaps 3;  
 QY 273 GTACGCCCGGCTTCTGCTGCGCACCCCTGCGCTGTGTACAGCGCGCGCAAGAGCGGTG 332  
 DB 2414838 GCGATCGCGGTCTCTGCGCGCAATCGCGCTACTGCGCGCGCGCGCGCGGTG 2414779  
 QY 333 CGCATCTCCGTGACATCATCTCTTCGCCG-----CGAGGCG 371  
 DB 2414778 CCGATCTGGGTGAGCTGGAGTTAGCTGGCGGTAGACGAGCGGTGCTACGAGCG 2414719  
 QY 372 AAGGCCCGGATCGTCCCATCATCACCGTTCCAAAGGAGAGAGACCGTGACCACTTCG 431  
 DB 2414718 CCGCGCAGCTGCTGTGTGTGACCGCGCACCAAGCAAGACACCGAGCTGATGCTG 2414659  
 QY 432 GCGCAATGCGGTGCGCGCGGACAAAGCGTGTGCGCGTGGCGGCAACCTCGCACCCCG 491  
 DB 2414658 CACGCGATGCTGATCGCGGTGCGCGCGCGGTGCTGTGCGCAATATCGCAGTGG 2414599  
 QY 492 GCGCTGACCTGCTGCGCGGACGACATCGAGCTGTAGTGTGCTGAGCTGTCCAG 551  
 DB 2414598 GTGCTGGATGCTGTGACGAGCGCGCGGAGCTGCTGCGGTGAGTGTTCAGTTCCAG 2414539  
 QY 552 CTGGAAACCTGCGATCGCTCAACCGCGAGGTGGCGGACCGTGTCTGAACCTCAGCGAAGAC 611  
 DB 2414538 CTGCACTGGCGCGCTGCTGCGCGCGCGGCGGTGCTCAACATTCGCCAAGAC 2414479  
 QY 612 CATATGATGCTGACGCGCATGCTGACTACCACTGGCGCAAGACCGGATCTTCGCG 671  
 DB 2414478 CACCTGGAGTGGCATGCGCAGCATGCGCAATACCGCGCGCAAGCGCGGTGCTGAC 2414419  
 QY 672 GGTGCGCGGAGTGTGCTGATGCGCGCGGTGCTGACCGCGCGGTGATCGCGCGGT 731  
 DB 2414418 GCGCGGGTAGCGGTGCGCGGTGCTGATGACAGCGCGCGGTGCTGCTGACGCGCTCA 2414359



QY 732 ACCGTGCGCTGCTGGTCG---TTCCGGCTGAAACAGCGGAGACTTCAAGGCTTTCCGGCTG 788  
 Db 2414358 CCGCGCAGAGTGCCTGGCTCGGCTTCGGCTCGCGAGCGCGCGGGGGAACACTGGCGTG 2414299  
 QY 789 ATCAGAGGAGACGCCAGAGTGGCTGGCTTCCAGTTCGACAGCTGCTGCGGTTGGC 848  
 Db 2414298 CGCAGCGCCACCTGGTGGCTTCCTCCGACGACTTGGACGCTGCTGCGGCTGCG 2414239  
 QY 849 GAACCTGAGATCCGTGGCGCCACACAACTATTCCAAACGCGCTTCGCGGCTGGCGTGGC 908  
 Db 2414238 TCGATACCGGTGCCAGGTCCGGTGGCTGCTTACGCGCTTGGCGCGCGGCTGGCC 2414179  
 QY 909 CATCGGCTGGCGCTTCCAGTTCGACGATGCTCGCGGCGCTGGAAGGCTTTTCGGGCTG 968  
 Db 2414178 CGCTGGTTCGGGCTGCGCGCGGCTGCGATCGCGGCTGCTGCTTTCGAGTGGC 2414119  
 QY 969 GCTCATGCTGCCAGTGGGTACGCGAGCGCGGAGGCGTGGCTACTACGAGATTCCAAG 1028  
 Db 2414118 CGACACCGCGGAGTGGTGGCGTTCGCCAGCGCATCTACGTGGAGCTCCAAG 2414059  
 QY 1029 GCCACCAACGTCGCGCGCCCTTGGCGGCGATCGAGGGGCTGGGTGCGGACATCGACGGC 1088  
 Db 2414058 GCCACCAACCGCGCGCGGCTTTCGGTGGCTG-----CATACCG 2414014  
 QY 1089 AAGCTGGTGTGCTCGCGCGGAGAGCGGAGGCGCGGATTTCCATGACCTCGCGGAG 1148  
 Db 2414013 AGGTTGGTATGATGCGCGGTGGCTGCTCAAGGCGCGTGGCTTCAAGCGGAGTTGCG 2413954  
 QY 1149 CCGGTGCGCGCTTCTGCGCGGCGGTGCTACTGCTTGGCGGTGACGCGGCGTATTGCG 1208  
 Db 2413953 GCGATGGCGTGGCGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 2413894  
 QY 1209 CAGGC 1213  
 Db 2413893 GAGGC 2413889

RESULT 13  
 ABZ41528  
 ID ABZ41528 standard; DNA; 1335 BP.  
 XX  
 AC ABZ41528;  
 XX  
 DT 07-MAR-2003 (first entry)  
 XX  
 DE N. gonorrhoeae nucleotide sequence SEQ ID 7645.  
 XX  
 KW Antibacterial; infection; vaccine; gene therapy; gene; ds.  
 XX  
 OS Neisseria gonorrhoeae.  
 XX  
 PN WO200279243-A2.  
 XX  
 PD 10-OCT-2002.  
 XX  
 PF 12-FEB-2002; 2002WO-IB02069.  
 XX  
 PR 12-FEB-2001; 2001GB-0003424.  
 XX  
 PA (CHIR-) CHIRON SPA.  
 XX  
 PI Fontana MR, Piazza M, Masignani V, Monaci E;  
 XX  
 DR WPI; 2003-058415/05.  
 DR P-PSDB; ABP80558.  
 XX  
 PT New protein from Neisseria gonorrhoeae, useful for the manufacture of a  
 PT medicament for treating or preventing N. gonorrhoeae infection -  
 XX  
 PS Disclosure; Page 745-746; 815pp; English.  
 XX  
 CC The present invention relates to proteins from Neisseria gonorrhoeae.  
 CC Also disclosed are the nucleic acid molecules encoding the proteins and

CC antibodies that specifically bind to the proteins. The composition  
 CC comprising the protein, nucleic acid or antibody is useful for the  
 CC manufacture of a medicament for treating or preventing N. gonorrhoeae  
 CC infection, this may be in the form of a vaccine or gene therapy.  
 CC Sequences given in records ABZ37706-ABZ42016 represent nucleic acid  
 CC molecules of the invention.  
 XX  
 SQ Sequence 1335 BP; 303 A; 341 C; 415 G; 276 T; 0 other;  
 Query Match 11.4%; Score 165.6; DB 25; Length 1335;  
 Best Local Similarity 49.08; Pred. No. 4.3e-21;  
 Matches 653; Conservative 0; Mismatches 629; Indels 50; Gaps 6;  
 QY 80 CATCGTGTGGCGCTCGCGAAGCGGCTATCCCTGGTGGCTACCTGGCGCGCGCG 139  
 Db 24 CCTCGTGGCGGACTTGGCGCGCGGCTATTTCCATGATTTCCCTATCTCGCAAAACGG 83  
 QY 140 CTTGCCCTTTCGCGTGTGCTATACCCGAGAGAACCCCGGAGTGGCCACCTCGCTGC 199  
 Db 84 CCGGAGGTTTGGCGGCTTATGATCGGAGCTGAAAGCGGAGCGGTGGCGCAAAATCGGTAA 143  
 QY 200 CCAGTATCCGAGGTGGAAGTGGCTTTCGCGGCAACTCGACGCGAGTTCTCTGCTCCGC 259  
 Db 144 GATGTTTACGGGCTGCTGTTTACAGGGTCTGCTGAAAGATGCGTTGGACACGGTTT 203  
 QY 260 CCGCGAACTCTATGTACGCCCCGGCTTGTGCTGCGCACCCCTGCTGTGTACAGGCGCG 319  
 Db 204 CGATATTCTGGCGCTAGTCCCGCATCAGCAGCGCGCGGATATCGAGGCGTTCAA 263  
 QY 320 CCGCAAGGCGTGGCATCTCCGCTGACATGCTTTCGCGCGCGAGCGGAGGCGCC 379  
 Db 264 GCAAAAGCGGCGCGCTGCTAGCGGACATCGAATGCTGGCGGACATTTGAAACCGCGC 323  
 QY 380 GATCGTC-----GCCATACCGGTTTCCAAAGCGGAGAGACCGCTGACCACTCGT 430  
 Db 324 CCGCGCAAGGTGATTCGGATTACCGCGCAGCAAGCGCAAAACCGGTACGAGCGTGGT 383  
 QY 431 GGGCGAAATGGCGGTGGCGCGGACAGCGTGTGCGCGTGGCGGCGAACCTCGCGCACCCC 490  
 Db 384 CCGCTATCTGCTCATCAAGTGGCGGCTGGATACCGTCTATTGCGGCGCAATATCGTAGCGC 443  
 QY 491 GCGCGTCG-----ACCTGCTGGCGGACGACATCGAGCTGTACCTGTGGAGCT 538  
 Db 444 GGTGTTGGAGGCGAGAAATTCGCGCGGAGGCAAAAGCGGCGAGCTGTGGGTGGAGCT 503  
 QY 539 GTCGAGCTTCCAGCTGGAAACCTTCCGATCGCTCAACGCGGAGGTGGCGACCGCTGCTGAA 598  
 Db 504 TTCAGCTTCCAACTGGAAACACCGGAAAGCGCTGGCGCGGAGTGGCGGATGATGAA 563  
 QY 599 CGTCAGCGAAGACCATATGATCGCTACGACGCGATGGCTGTACCATCGCTGCCACGCA 658  
 Db 564 CATTTCCGAAGACCATCTCGACCGCTACGACGATTTGCTCGACTACGCGCATACCAAAGC 623  
 QY 659 CCGGATCTCCCGGCTGGCGCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 718  
 Db 624 CGAGATTTTCGCTGGCGGCTGAGTGGAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 683  
 QY 719 GCTG-ATCGCGGATACCGTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 777  
 Db 684 CATGAACGGGCGAGGCGGTGAAGTGAACCGGTTTTCCTTGGAAACAGAGCCGATTTTG 743  
 QY 778 CTTTCGCGCTGATCGAGGAGAGCGGCGAGAGTGGCTGGCGTTTCCAGTTTCGACAAGCTGC 837  
 Db 744 GTT-----GGAACGCGGCGCGGCTGTTTAAACAAGCAATGAAGATTGA 790  
 QY 838 TGGCGGTTGCGCAACTGAAGATCCGTGGCGCGCCACAACTATTCCACAGCGCTCGCGCGC 897  
 Db 791 TTTTCTACGCAAGACATCCCGCTGCAAGGTTTGGACAAACCGCGCGCAACGTTATGGCTGCC 850  
 QY 898 TGGCGCTGGCGCATCGGCTGGCGCTTCCGCTTCCGACCGCATGCTGCGCGCGCTGAAGCGCT 957  
 Db 851 TGGCTTTTGGAGGCGGCTGGTTTGGCGGCGGAGCAATGCTGGAAACAGCTGCAAAACCT 910



```
QY 958 TTTCCGGCTGCTCATCGTCCAGTGGGTACGGAGCGCGAGCGGTGAGCTACTACG 1017
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 911 TCCAAGGCTTGGCCACCGCTGTGGAATAATCGCGAGAAACCGCGTGTATTCATCG 970
QY 1018 AGATTCCAAAGCCACCAACGTGCGCGCGCGCCCTCGCGCGGATGAGGGGCTGGCGG 1077
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 971 ACACAGCAAAAGGACGAAATGTCGGCGGACCGCGCGCGAT-----TGCGG 1018
QY 1078 ACATCGACGGCAAGCTGTGCTCGCGCGCGGAGACGGCAAGGCGCGGATTTCCATG 1137
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1019 GTTTCGAACCCGCTCTCTGATTTGGCGGCGATGGCCAAAGGAGGACTTCACGC 1078
QY 1138 ACTCGCGAGCGGTGCGCGCTTCTGCGCGCGGTGTACTGTGCGCGGTGACGCGG 1197
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1079 CCCTCGCGACCGTTGAAGATAAGGCAAAAGCGGTGCTCTGATCGCGTCGATCGCG 1138
QY 1198 GCGTATGTCACGACATG---GGCAAGCGGTACCGTGTGCGCGGTGCGCAAGCTGG 1254
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1139 CGCAATCGCGCGGATTTGGACGGCTGCGGCTTGAACCTTGACCGACTGCGTCACTTTGG 1198
QY 1255 ACGAAGCAGTCCGCGCGCGCGAGCTGGCGCGGAGCGCAAGGCGATCGGTGCTGTCGC 1314
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1199 AGAGCGGTTGACAGCGCATACGCCCAAGCGGCGGATATTTGCTGCTCAGCG 1258
QY 1315 CGGCTGCGCGAGCTGGACATGTTCAAGAACTTCGAAGAACGCGGACGCGCTGTTGCGCA 1374
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1259 CGGCTGCGCGAGTTTCGATATGTTTAAAGGCTACGCGCACCGGTTTCGGAAGTGTATCG 1318
QY 1375 AAGCGTAGAGG 1386
  ||| ||| |||
Db 1319 AAGCGTTAAGG 1330
  ||| ||| |||
```

RESULT 14

```
AAS89816
ID AAS89816 standard; cDNA; 4942 BP.
XX
AC AAS89816;
XX
DT 13-FEB-2002 (first entry)
XX
DE DNA encoding novel human diagnostic protein #25620.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
XX
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI; 2001-639362/73.
XX
DR P-PSDB; ABG25629.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
PS Claim 1; SEQ ID No 25620; 103pp; English.
XX
PS
XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
```

```
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags.
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: the sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 4942 BP; 1109 A; 1296 C; 1387 G; 1150 T; 0 other;
Query Match 10.9%; Score 158; DB 23; Length 4942;
Best Local Similarity 57.1%; Pred. No. 1e-19;
Matches 312; Conservative 0; Mismatches 225; Indels 9; Gaps 1;
QY 80 CATCGTTGTCGCGCTCGCAAGAGCGGCATGTCCTGTGGTGCCTACTCGCGCGCGCG 139
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1749 CGTCATATTCGCGCTGGGCTTCACGGGCTTCTTCGTGGTGGACTTTTCTCGCTCGCG 1808
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 140 CTTGCTTTTCGCGGTGTCGATACCCGAGAGAACCCGCGGAGCTGGCCACCTCGGTGC 199
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1809 TGTGACGCGCGCGTTATGATACGCTATACACCGCTGCTGGATAAATACCGC- 1867
QY 200 CCAGTATCCGAGGTGGAAGTGGTGGCGGAACTCGACGCGGAGTTCCTCTGCTCCGC 259
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1868 -----AAGCGGTAGAACGCCACACGGGCGAGTCTGAATGATGCTGATGGCGGC 1919
QY 260 CGCGAACTCTATGTCAGCCCGCGGCTTGTGCTGCGCACCCCTGCGCTGTACAGGCGCG 319
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1920 AGATCTGATGTCGCCAGTCCCGGTATTGCACTGGCGCATCCATCCTTAAGCGCTGCCGC 1979
QY 320 CGCGAAAGCGGTGCGCATCTCCGGTGACATCGATCTCTTTCGCGCGGAGGCGAAGCGCC 379
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1980 TGATCGCGGAATCGAAATCGTTGGCGATATCGAGCTGTCTGTGCGAAGCACAAGCAC 2039
QY 380 GATCGTCGCCATCAGCGGTTCCACGCGAAGAGCACCGTGACCCCTCTGTGGCGAAT 439
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2040 GATTGTGGGATTACCGGTTCTAAGGGCAAAAGCAGGTCACCCAGTGTGGGTGAAT 2099
QY 440 GCGGTGGCGCGGACAAGCGTGTGCGCCCTGCGCGGCAACCTCGGCACCCCGCGCTCGA 499
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2100 GCGAAAGCGCGCGGGGTTAACGTTGGTGTGGTGGCAATATGGCTGCTGCTGAT 2159
QY 500 CCGTGTGCGCGGACGATCGAGCTGTACGTTGGAGCTGTGAGCTTCCAGCTGGAAC 559
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2160 GCTACTGGATGATGATGTGAATCTGCTGCTGGAACCTGTGAGCTTCCAGCTGGAAC 2219
QY 560 CTGCGATCGCTCAACGCGGAGGTGGCGACCTGTGTAACCTCAGCGAAGACCATATGA 619
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2220 CACCTCCAGCTTACAGGCGGTAGCAGCGACCATTTCTGAACGCTGACTGATATGGA 2279
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 620 TCGCTA 625
  ||| ||| |||
Db 2280 TCGCTA 2285
  ||| ||| |||
RESULT 15
AAS89273
ID AAS89273 standard; cDNA; 4944 BP.
XX
XX AAS89273;
AC AAS89273;
XX
```

DT 13-FEB-2002 (first entry)  
DE DNA encoding novel human diagnostic protein #25077.  
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.  
XX Homo sapiens.  
OS  
XX  
PN WO200175067-A2.  
XX  
XX 11-OCT-2001.  
PD  
XX  
XX 30-MAR-2001; 2001WO-US08631.  
PF  
XX 31-MAR-2000; 2000US-0540217.  
PR 23-AUG-2000; 2000US-0649167.  
XX  
XX (HYSE-) HYSEQ INC.  
PA  
XX Drmanac RT, Liu C, Tang YT;  
PI  
XX P-PSDB; ABG25086.  
DR  
XX  
XX New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity -  
XX  
PS Claim 1; SEQ ID No 25077; 103pp; English.  
XX  
XX The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. AAS64197-AAS94564 represent novel human  
CC diagnostic coding sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 4944 BP; 1110 A; 1302 C; 1383 G; 1149 T; 0 other;

Query Match 10.9%; Score 158; DB 23; Length 4944;  
Best Local Similarity 57.1%; Pred. NO. 1e-19;  
Matches 312; Conservative 0; Mismatches 225; Indels 9; Gaps 1;

QY 80 CATCGTTGTCGGCTCGCAGAGCGCATCTCCCTGGTGGCTACCTGGCGCCCGCGG-139  
DB 1749 CGTCATTATCGGCTGGGCTCCACGGGCTTTCTCGTGGACCTTTTCTCGCTCGCGG 1808  
QY 140 CTTGCCCTTTCCGCTGGTTCGATACCCGAGAGAACCCGCGGAGCTGGCCACCTGCGTGC 199  
DB 1809 TGTGACGCCGCGCTTATGGATACCGCTATGACACCGCTGGCTGGATAAATTACCG- 1867  
QY 200 CCAGTATCCGAGGTGGAAGTGGCTTCGGGGGAATCTGACGCCGAGTTCCTGCTCCGC 259  
DB 1868 -----AAGCCGTAGAACGCCACACGGGCGAGTCTGAATGATGAATGGCTGATGGCGG 1919  
QY 260 CCGGGAACCTCTATGTCAGCCCCGGCTTGTGCTCGCGCACCCCTCGCTGTACAGGCCG 319

DB 1920 AGATCTGATTGTCGCCAGTCCCGGTATTGCACTGGCGCATCCATCTTAAGCGCTGCCGC 1979  
QY 320 CGCGAAAGGCGTGGCGATCTCCGGTGGACATCTCTTCGCCGCGAGGCGAAGGCCCC 379  
DB 1980 TGATGCCGGAATCGAATTCGTTGGGATATCGAGCTGTTCTGCCGAAGCACAAGCACC 2039  
QY 380 GATCGTCGCCATCACCGGTTCCAAACGCGAAGAGCACCGTGACCAACCTTGGTGGGCGAAAT 439  
DB 2040 GATTGTGGCGATTACCGGTTCTAAGGGCAAAAGCACGCTACACGCTAGTGGGTGAAT 2099  
QY 440 GCGGTGGCCGCGGACAAAGCGTGTGCGCGTGGGGGCAACCTCGGCAACCCCGGCGCTCGA 499  
DB 2100 GCGAAAGCGCGGGGTTAACTGTTGGTGGTGGCAATATTGGCTGCGCTGCGCTTGTAT 2159  
QY 500 CCTGCTGGCCGACGACATCGAGCTGTAGTGTGTTGAGCTGTGAGCTTCCAGCTGGAAC 559  
DB 2160 GCTACTGGATGATGAGTGTGAACGTGTAGTGTGGAACGTGTGAGCTTCCAGCTGGAAC 2219  
QY 560 CTGCGATCGCCTCAACGCGAGGTGGCGACCGTGTGAACGTGACGAGAACCACTATATGA 619  
DB 2220 CACCTCCAGCTTACAGGCGGTAGCAGCGACCATTTCTGAACGTGACTGAAGATCATATGA 2279  
QY 620 TCGCTA 625  
DB 2280 TCGCTA 2285

Search completed: August 14, 2003, 01:55:30  
Job time : 443 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: August 14, 2003, 07:06:25 ; Search time 92 Seconds  
(without alignments)  
2149.344 Million cell updates/sec

Title: US-09-701-229-2

Perfect score: 2260

Sequence:

1 MSIIASHDFRIVVGLKSGM.....MFKNFERGRFLAKAVEELA 448

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 712578

Minimum DB seq length: 25

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+\_p2n.model -DEV=xlp  
-O=/cgn2\_1/USPTO.spool/US09701229/runat\_09082003.161158.4405/app\_query.fasta.1.647  
-DB-Issued\_Patents\_NA\_QFMT-fastop -SUFFIX=p2n.rn1 -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS-human40.cdi  
-LIST=45 -DOCALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=25 -MAXLEN=2000000000  
-USER=US09701229\_QCGN\_1.1.85 -runat\_09082003.161158.4405 -NCPU=6 -ICPU=3  
-NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Issued\_Patents\_NA.\*  
1: /cgn2\_6/ptodata/1/ina/5A\_COMB.seq.\*  
2: /cgn2\_6/ptodata/1/ina/5B\_COMB.seq.\*  
3: /cgn2\_6/ptodata/1/ina/6A\_COMB.seq.\*  
4: /cgn2\_6/ptodata/1/ina/6B\_COMB.seq.\*  
5: /cgn2\_6/ptodata/1/ina/PTUS\_COMB.seq.\*  
6: /cgn2\_6/ptodata/1/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2260	100.0	1371	4	US-09-252-991A-7702
c 2	2252	99.6	1401	4	US-09-252-991A-7861
3	1148.5	50.8	1416	4	US-09-328-352-1282
c 4	951	42.1	567	4	US-09-252-991A-7787
5	879	38.9	1830121	4	US-09-557-884-1
6	879	38.9	1830121	4	US-09-643-990A-1
c 7	626	27.7	564	4	US-09-252-991A-7928
8	603.5	26.7	2193	2	US-08-934-481-1
9	603.5	26.7	2193	4	US-09-290-602-1
c 10	594.5	26.3	6693	4	US-08-961-527-195
11	591	26.2	360	4	US-09-252-991A-7552
12	586.5	26.0	1350	1	US-08-665-435A-1

13	586.5	26.0	1350	1	US-08-665-435A-3
14	586.5	26.0	1350	2	US-08-843-309-3
15	586.5	26.0	1353	2	US-08-843-309-1
c 16	525	23.2	4411529	3	US-09-103-840A-1
c 17	515	22.8	4403765	3	US-09-103-840A-2
18	504.5	22.3	1368	4	US-09-107-532A-1574
c 19	493	21.8	1003	3	US-08-961-083-109
20	493	21.8	1003	4	US-09-536-784-109
21	470	20.8	1372	4	US-09-530-836-1
22	469	20.8	1374	4	US-09-134-001C-1001
23	421	18.6	1242	3	US-09-147-928-1
24	395	17.5	2423	3	US-08-714-918-86
25	395	17.5	2423	3	US-09-265-315-86
26	395	17.5	2423	3	US-09-265-315-86
27	395	17.5	2423	3	US-09-266-417-86
28	391.5	17.3	1230025	4	US-09-198-452A-1
29	389.5	17.2	2402	4	US-09-221-017B-881
30	379	16.8	801	2	US-08-934-481-3
31	379	16.8	801	4	US-09-290-602-3
32	321	14.2	1359	4	US-09-252-991A-7624
33	321	14.2	2393	3	US-09-230-380-8
c 34	292	12.9	843	4	US-08-961-527-356
35	240	10.6	1515	4	US-09-252-991A-7701
c 36	210.5	9.3	1215	4	US-09-252-991A-7862
37	207	9.2	1360	2	US-08-732-612-1
38	195.5	8.7	1479	4	US-09-328-352-897
39	195.5	8.7	1542	4	US-09-252-991A-7625
c 40	195.5	8.7	1701	4	US-09-252-991A-7927
41	195	8.6	1503	4	US-09-328-352-2140
42	182.5	8.1	1717	4	US-08-936-165A-78
43	182.5	8.1	1452	4	US-09-328-352-1894
44	173	7.7	536	3	US-09-147-928-3
45	170.5	7.5	1400	3	US-09-144-918-1

#### ALIGNMENTS

##### RESULT 1

US-09-252-991A-7702  
; Sequence 7702, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 7702  
; LENGTH: 1371  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-7702

Alignment Scores:  
Pred. No.: 3.5e-241 Length: 1371  
Score: 2260.00 Matches: 448  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 4 Gaps: 0

US-09-701-229-2 (1-448) x US-09-252-991A-7702 (1-1371)

Qy 1 MetSerLeuIleAlaSerAspHisPheArqIleValValcIlyLeuGlyLysSerGlyMet 20  
|||||  
Db 25 ATGAGCCTGATGCCTCCGACCACCTCCCGATCGTGTTCGCGCAAGCGGCATG 84

```
QY 21 SerLeuValArgTyrLeuAlaArgArgGlyLeuProPheAlaValAlaAspThrArgGlu 40
Db 85 TCCTTGGTGGCTACCTGGCGCGCGCGCTTTCCTTTTCGCCGTTGTGCATACCGAGAG 144
QY 41 AsnProProGluLeuAlaThrLeuArgAlaGlnTyrProGlnValGluValArgCysGly 60
Db 145 AACCCGCGGAGCTGGCCACCTTGGTGCCAGATATCCGACAGTGGAAAGTGGCTTTCGCGC 204
QY 61 GluLeuAspAlaGluPheLeuCysSerAlaArgGluLeuTyrValSerProGlyLeuSer 80
Db 205 GAACCTGACGCGGAGTTCTCTGTCTCCGCCCGGAACCTCTACGTCAGCCCGGCTTGTGC 264
QY 81 LeuArgThrProAlaLeuValGlnAlaAlaLysGlyValArgIleSerGlyAspIle 100
Db 265 CTGCGCACCCCGCGCTGTGTACAGCGCGCGGAAAGTGTGCGATCTCCGGTGACATC 324
QY 101 AspLeuPheAlaArgGluAlaLysAlaProIleValAlaIleThrGlySerAsnAlaLys 120
Db 325 GATCTCTTCCCGCGGAGCGAGGCCCGCGATGTCGCCATCACCCTGTTCCAACCGGAG 384
QY 121 SerThrValThrLeuValGluGluMetAlaValAlaAlaAspLysArgValAlaVal 140
Db 385 AGCACCGTGACCACTGGTGCGGAAATGGCGGTGGCGCGACAAAGCGTGTCCGCGTC 444
QY 141 GlyGlyAsnLeuGlyThrProAlaLeuAspLeuLeuAlaAspIleGluLeuTyrVal 160
Db 445 GCGCGCAACTCGGCACCCCGCGCTCGACCTGCTGGCGGACGATCGAGTGTACGTG 504
QY 161 LeuGluLeuSerSerPheGlnLeuGluThrCysAspArgLeuAsnAlaGluValAlaThr 180
Db 505 TTGGAGCTGTCCAGCTTCCAGCTGGAACCTCGCATCGCTCAACCGGAGTGGCGACC 564
QY 181 ValLeuAsnValSerGluAspHisMetAspArgTyrAspGlyMetAlaAspTyrHisLeu 200
Db 565 GTGCTGAACGTCAGCGAAGACCATATGGATGCTAGCAGCGCATGCTACTACCACTG 624
QY 201 AlaLysHisArgIlePheArgGlyAlaArgGlnValValAlaAsnArgAlaAspAlaLeu 220
Db 625 GCCAAGCACCGGATCTTCGCGGTGCGCGCCAGTCTGGTGAATCGCGCGCATGCCCTG 684
QY 221 ThrArgProLeuIleAlaAspThrValProCysTrpSerPheGlyLeuAsnLysProAsp 240
Db 685 ACCCGACCGCTGATCGCGATACCGTGGCTGCTGTGTCGGCTGTAACAAAGCCGAC 744
QY 241 PheLysAlaPheGlyLeuIleGluAspGlyGlnLysTrpLeuAlaPheGlnPheAsp 260
Db 745 TTCAAGGCTTTCGGCTGATCAGGAGAGAGCCAGAGTGGCTGGCGTTCCAGTTCCAG 804
QY 261 LysLeuLeuProValGlyGluLeuLysIleArgGlyAlaHisAsnTyrSerAsnAlaLeu 280
Db 805 AAGCTGCTGCGCGTTGGCGAATGAAGATCCGTGGCGCCCACTATTTCCAACGCGCTC 864
QY 281 AlaAlaLeuAlaLeuGlyHisAlaValGlyLeuProPheAspAlaMetLeuGlyAlaLeu 300
Db 865 GCGCGCTGGCGTGGCCATGCGGTGCGGCTGCGGTTCGACGCCATGCTGCGCGCGCTG 924
QY 301 LysAlaPheSerGlyLeuAlaHisArgCysGlnTrpValArgGluArgGlnGlyValSer 320
Db 925 AAGCGGTTTTCGGCTGCTGCTCATGCTGCCAGTGGGTACGCGAGCGCGAGGGCGTGC 984
QY 321 TyrTyrAspAspSerLysAlaThrAsnValGlyAlaAlaLeuAlaAlaIleGluGlyLeu 340
Db 985 TACTACGACGATTTCAAGGCCACCAACGTGCGCGCGCGCTTGGCGGCGATCGAGGGGCTG 1044
QY 341 GlyAlaAspIleAspGlyLysLeuValLeuLeuAlaGlyLysGlyValAlaAsp 360
Db 1045 GGTGCGGACATCGAGGCAAGCTGTGTGCTGCGCGCGGAGACGGCAAGGGCGCGAT 1104
QY 361 PheHisAspLeuArgGluProValAlaAlaArgPheCysArgAlaValValLeuLeuGlyArg 380
Db 1105 TTCCATGACCTCGCGGAGCGGTGCGCGCTTCTGCGCGGCGGTGCTACTGCTTGGCGGT 1164
QY 381 AspAlaGlyLeuIleAlaGlnAlaLeuGlyAsnAlaValProLeuValValArgValAlaThr 400
```

```
Db 1165 GACGCGGCGCTGATTGCCAGGCACTGGGCAACGCGGTACCGCTGGTGGCGTCGCAACG 1224
QY 401 LeuAspGluAlaValArgGlnAlaAlaGluLeuAlaArgGluGlyAspAlaValLeuLeu 420
Db 1225 CTGGACGAGACGAGTCCGCGAGCCCGCGAGCTGGCCCGGAAGGCGATGCGGTGTGTG 1284
QY 421 SerProAlaCysAlaSerLeuAspMetPheLysAsnPheGluGluArgGlyArgLeuPhe 440
Db 1285 TCGCGCGCTGCGCGAGCTGGACATGTTCAAGAACTTCGAAGAACGCGGAGCGCTGTC 1344
QY 441 AlaLysAlaValGluGluLeuAla 448
Db 1345 GCCAAAGCGGTAGAGGAGCTAGCG 1368

RESULT 2
US-09-252-991A-7861/c
; Sequence 7861, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ IDS: 33142
; SEQ ID NO 7861
; LENGTH: 1401
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-7861

Alignment Scores:
Pred. NO.: 2,8e-240 Length: 1401
Score: 2252.00 Matches: 446
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 99.65% Indels: 0
DB: 4 Gaps: 0

US-09-701-229-2 (1-448) x US-09-252-991A-7861 (1-1401)
QY 1 MetSerLeuIleAlaSerAspHisPheArgIleValValGlyLeuGlyLysSerGlyMet 20
Db 1338 ATGAGCCTGATCGGCTCCGACCACTTCGCGCATCGTTGCGGCCCTCGGCAAGCGGCATG 1279
QY 21 SerLeuValArgTyrLeuAlaArgGlyLeuProPheAlaValAlaAspThrArgGlu 40
Db 1278 TCCCTGGTGGCTACCTGGCGCGCGCGCTTGGCTTTCGCGCTGTGTGCATACCGGAG 1219
QY 41 AsnProProGluLeuAlaThrLeuArgAlaGlnTyrProGlnValGluValArgCysGly 60
Db 1218 AACCGCGGAGCTGGCCACCTGCGTCCCGCATATCGCAGGTGGAAGTGGCTTTCGCGC 1159
QY 61 GluLeuAspAlaGluPheLeuCysSerAlaArgGluLeuTyrValSerProGlyLeuSer 80
Db 1158 GAACCTGACGCGGAGTTCTCTGCTCGCGCGCGCAACTCTACGTACGCGCGCGGTGTCG 1099
QY 81 LeuArgThrProAlaLeuValGlnAlaAlaLysGlyValArgIleSerGlyAspIle 100
Db 1098 CTGGCACCCCGCGCTGGTACAGCGCGCGCGAAGGTTGCGCATCTCCGGTGACATC 1039
QY 101 AspLeuPheAlaArgGluAlaLysAlaProIleValAlaIleThrGlySerAsnAlaLys 120
Db 1038 GATCTTTCGCGCGGAGCGGAGGCCCGCATGCTCGCCATCACCGGTTCACACGCGAAG 979
QY 121 SerThrValThrLeuValGlyGluMetAlaValAlaAspLysArgValAlaVal 140
```

Db 978 AGCACCTGACCAACCTGGTGGCGAAATGCGGTGGCGGACAAAGCGTGTGCGCGTC 919  
Qy 141 GlyGlyAsnLeuGlyThrProAlaLeuAspLeuAlaAspAspIleGluLeuTyrVal 160  
Db 918 GCGGCAACCTCGGCAACCGCGCTGACCTGCTGGCGACGACATCGAGCTGACGTG 859  
Qy 161 LeuGluLeuSerSerPheGlnLeuGluThrCysAspArgLeuAsnAlaGluValAlaThr 180  
Db 858 TTGGAGCTGTGAGCTCCAGCTGGAACCTGCGCTCAACGCGGAGGTGGCGACC 799  
Qy 181 ValLeuAsnValSerGluAspHisMetAspArgTyrAspGlyMetAlaAspTyrHisLeu 200  
Db 798 GTGTGACGTCACGCAAGACCATATGATGCTGACGACGCGATGGCTGACTACCACTG 739  
Qy 201 AlalysHisArgIlePheArgGlyAlaArgGlnValValAlaAsnArgAlaAspAlaLeu 220  
Db 738 GCCAAGCACCGGATCTCCGGGTGCCCGCAGAGTGTGTGAATCGCGCGATGCCCTG 679  
Qy 221 ThrArgProLeuIleAlaAspThrValProCysTrpSerPheGlyLeuAsnLysProAsp 240  
Db 678 ACCGAGCGCTGATCGCGATACCGTGGCTGGTGGTGGCGCTGAACAACGCCGAC 619  
Qy 241 PheLysAlaPheGlyLeuIleGluAspGlyGlnLysTrpLeuAlaPheGlnPheAsp 260  
Db 618 TTCAAGGCTTTCCGGCTGATCGAGAAAGACGGCAGAGTGGCTGGCGCTTCCAGTTCCGAC 559  
Qy 261 LysLeuLeuProValGlyGluLeuLysIleArgGlyAlaHisAsnTyrSerAsnAlaLeu 280  
Db 558 AAGCTGTGCGGTTGGGAACTGAAGTCGTGGCGGCCACACTATTCACAGCGCTC 499  
Qy 281 AlaAlaLeuAlaLeuGlyHisAlaValGlyLeuProPheAspAlaMetLeuGlyAlaLeu 300  
Db 498 GCCGCGTGGCGCTGGGCGATCGCGTGGCGTGGCGTGGCGTGGCGTGGCGTGGCG 439  
Qy 301 LysAlaPheSerGlyLeuAlaHisArgCysGlnTrpValArgGluArgGlnGlyValSer 320  
Db 438 AAGCGTTTTCCGGCTGGCTATCGCTGCGCGTGGGTACCGAGCGCGCGCGTGGCG 379  
Qy 321 TyrTyrAspSerLysAlaThrAsnValGlyAlaAlaLeuAlaIleGluGlyLeu 340  
Db 378 TACTACGACGATTCACAGGCCACCAACGTCGCGGCCGCCCTGGCGGCGATCGAGGCGT 319  
Qy 341 GlyAlaAspIleAspGlyLysLeuValLeuLeuAlaGlyLysGlyLysGlyAlaAsp 360  
Db 318 GGTCCGACATCGACGCAAGCTGCTGCTCGCGCGGAGACGCGCAAGGCGCGCAT 259  
Qy 361 PheHisAspLeuArgGluProValAlaArgPheCysArgAlaValAlaLeuGlyArg 380  
Db 258 TTCCATGACCTGCGGAGCGCGTCTGCGCGCTTCTGCGGCGCGGTGACTGCTGGCGGT 199  
Qy 381 AspAlaGlyLeuIleAlaGlnAlaLeuGlyAsnAlaValProLeuValArgValAlaThr 400  
Db 198 GACGCGGCGTGAATGCCAGGCACCTGGGCAACGCGTACCGCTGGTGGCGCGCAACG 139  
Qy 401 LeuAspGluAlaValArgGlnAlaGluLeuAlaArgGluGlyAspAlaValLeuLeu 420  
Db 138 CTGGACGAAGCAGTCCGCGAGCGCGCGAGTGGCGCGCGAAGCGATGCGGCTGCTGTG 79  
Qy 421 SerProAlaCysAlaSerLeuAspMetPheLysAsnPheGluGluArgGlyArgLeuPhe 440  
Db 78 TCGCGCGCTGCGGAGCGCTGGAGATGTTCAAGAACTTCGAAGAAGCGCGGAGCGCTGTC 19  
Qy 441 AlalysAlaValGluGlu 446  
Db 18 GCCAAGCGCTAGAGGAG 1

## RESULT 3

US-09-328-352-1282  
; Sequence 1282, Application US/09328352  
; Patent No. 6562958  
; GENERAL INFORMATION:  
; APPLICANT: Gary L. Breton et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER

; TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: GTC99-03PA  
; CURRENT APPLICATION NUMBER: US/09/328,352  
; CURRENT FILING DATE: 1999-06-04  
; NUMBER OF SEQ ID NOS: 8252  
; SEQ ID NO 1282  
; LENGTH: 1416  
; TYPE: DNA  
; ORGANISM: Acinetobacter baumannii  
US-09-328-352-1282

Alignment Scores:  
Pred. No.: 8,45e-118 Length: 1416  
Score: 1148.50 Matches: 240  
Percent Similarity: 66.44% Conservative: 59  
Best Local Similarity: 53.33% Mismatches: 142  
Query Match: 50.82% Indels: 9  
DB: 4 Gaps: 4  
US-09-701-229-2 (1-448) x US-09-328-352-1282 (1-1416)  
Qy 3 LeuIleAlaSerAspHisPheArgIleValValGlyLeuGlyLysSerGlyMetSerLeu 22  
Db 73 TTAATACACGTGGTGGATTAAAGTTGTAGCAGGCTGGGAATATCAGGTGTTCTGCT 132  
Qy 23 ValArgTyrLeuAlaArgArgGlyLeuProPheAlaValValAspThrArgGluAsnPro 42  
Db 133 GTAATTTCTTCGTCATGAACAGGCTACCAAGTCTGTAAACGATTCCTGGTGGTCA 192  
Qy 43 ProGluLeuAlaThrLeuArgAlaGlnTyrProGln---ValGluValArgCysGlyGlu 61  
Db 193 CCCGGA-----CACGATCAGATTCGGCTGGTGTAAACACGATTTTGGTCAG 240  
Qy 62 LeuAspAlaGluPheLeuCysSerAlaArgGluLeuTyrValSerProGlyLeuSerLeu 81  
Db 241 CTGATCAAGAAATATTATTACAGCAGAGAGATTTTAAAGCCAGGCTTGCACCA 300  
Qy 82 ArgThrProAlaLeuValGlnAlaAlaLysGlyValArgIleSerGlyAspIleAsp 101  
Db 301 CAATTACCGAAATTCAGGCAGCTATTCTCTAAAGTATTCTGTGGTGGCGATATCCAG 360  
Qy 102 LeuPheAlaArgGluAlaLysAlaProIleValAlaIleThrGlySerAsnAlaLysSer 121  
Db 361 TTATTTGCGCGTCCACTGATGTCGCGATTCGCGATTCAGAGTTCCAAATGCAAAAAGT 420  
Qy 122 ThrValThrThrLeuValGlyGluMetAlaValAlaAlaAspLysArgValAlaValGly 141  
Db 421 ACTGTACCACTTTTAATTTGGCTTAATGGCTAAGGATGCGAGTAAGAAAGTTGCAATAGGC 480  
Qy 142 GlyAsnLeuGlyThrProAlaLeuAspLeuAlaAspAspIleGluLeuTyrValLeu 161  
Db 481 GGCAACCTTGGTGCACGACGCTTAGATTTACTTAAAGATCAACACGAGATTACTGGTCTCT 540  
Qy 162 GluLeuSerSerPheGlnLeuGluThrCysAspArgLeuAsnAlaGluValAlaThrVal 181  
Db 541 GAGTTATCAAGCTTTCAGTTTAGAAACCACTCTCCTCAATAAACGCTGAGGTAGCAGTGGT 600  
Qy 182 LeuAsnValSerGluAspHisMetAspArgTyrAspGlyMetAlaAspTyrHisLeuAla 201  
Db 601 CTCAATATGATGAAGACCAATTTAGCGCTCATGGAATATGCTGGGGTATCATCAAGCA 660  
Qy 202 LysHisArgIlePheArgGlyAlaArgGlnValValValAsnArgAlaAspAlaLeuThr 221  
Db 661 AACATCGTATTTTCCAGGCGCTAANAAGTTGTATTAAACGAGATGATGCGTGAAGC 720  
Qy 222 ArgProLeuIleAlaAspThrValProCysTrpSerPheGlyLeuAsnLysProAspPhe 241  
Db 721 CGTCCACTTGTTCAGATACGACCAATGCAAGCTTTGCTTTTAAACGACCGCGATTTA 780  
Qy 242 LysAlaPheGlyLeuIleGluGlu---AspGlyGlnLysTrpLeuAlaPheGlnPheAsp 260  
Db 781 AATCAATATGCGCGTTTAAAGATGCGGACGCTAGCTTTGGCTTGTGCTGTTTACAG 840

QY 261 LysLeuLeuProValGlyGluLeuLysIleArgGlyAlaHisAsnTyrSerAsnAlaLeu 280  
Db 841 CGGTTAAATAAAGCTCAGATTATATATCAAGATGATGCACAACTAGTAAATGCTTTA 900  
QY 281 AlaAlaLeuAlaLeuGlyHisAlaValGlyLeuProPheAspAlaMetLeuGlyAlaLeu 300  
Db 901 GCTTGTTAGCATGGGGGAAGCAATGGTTTACCTATGAGTCAATGCTTGAACCTTTA 960  
QY 301 LysAlaPheSerGlyLeuAlaHisArgCysGlnTrpValArgGluArgGlnGlyValSer 320  
Db 961 AAAAAATTTAAAGCAGTAGACACCCCTGTGAGTACGTTAAACCCGTGATGTCCT 1020  
QY 321 TyrTyrAspAspSerLysAlaThrAsnValGlyAlaAlaLeuAlaAlaLeuGlyLeu 340  
Db 1021 TATTATAATGACTCTAAAGGTACTAATGTTGGTCTACACTTCAGCAATTTGATGCTTA 1080  
QY 341 GlyAlaAspLeuAsp-----GlyLysLeuValLeuAlaGlyGlyAspGlyLys 357  
Db 1081 GGTGCTGCCATTAAGTGAAGTAAGGTGACCTTATTTAGTGGACAGGCAAA 1140  
QY 358 GlyAlaAspPheHisLeuArgGluProValAlaAlaArgPheCysArgAlaValAlaLeu 377  
Db 1141 GGACAAGATTTTCTCTTACGTTCTCTATCGAAATATATCCCAAGTTGTGATTTG 1200  
QY 378 LeuGlyArgAspAlaGlyLeuIleAlaGlnAlaLeuGlyAsnAlaValProLeuValArg 397  
Db 1201 ATTGGTGAAGATGCGCTGCTCATCGAACAGCCATTCAGAGGCGCACTAAATTTTACAT 1260  
QY 398 ValAlaThrLeuAspGluAlaValArgGlnAlaAlaGluLeuAlaArgGluGlyAspAla 417  
Db 1261 GCAGCAACGCTTAAAGAACGCTGCGAGCTGTGTCACGCTGAACACCAAGCTGAAGATGTG 1320  
QY 418 ValLeuLeuSerProAlaCysAlaSerLeuAspMetPheLysAsnPheGluGluArgGly 437  
Db 1321 GTATTGCTATCACCAGCATGTGCAAGCTTTTGATATGTTTAAAGTTTATAATGACCGTGT 1380  
QY 438 ArgLeuPheAlaLysAlaValGluGluLeu 447  
Db 1381 CAGCAGTTTGTGCTGCGTCAATTCGTTG 1410

RESULT 4  
US-09-252-991A-7787/c  
; Sequence 7787, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 7787  
; LENGTH: 567  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-7787

Alignment Scores:  
Pred. No.: 1,72e-96 Length: 567  
Score: 951.00 Matches: 188  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 42.08% Indels: 0  
DB: 4 Gaps: 0

US-09-701-229-2 (1-448) x US-09-252-991A-7787 (1-567)

QY 242 LysAlaPheGlyLeuIleGluGluAspGlyGlnLysTrpLeuAlaPheGlnPheAspLys 261

Db 565 AAGGCTTTGGCGCTGATCGAGGAAGACGCCAGAGTGGCTGGCTTCCAGTTCGACAAG 506  
QY 262 LeuLeuProValGlyGluLeuLysIleArgGlyAlaHisAsnTyrSerAsnAlaLeuAla 281  
Db 505 CTGCTCCCGGTTGGCAACTGAAGATCCGTGGCGCCCACTATTTCACACGCGCTCGCC 446  
QY 282 AlaLeuAlaLeuGlyHisAlaValGlyLeuProPheAspAlaMetLeuGlyAlaLeuLys 301  
Db 445 GCGCTGGCGCTGGCCATCGGTCGCGCTGCCGTCGACGCCATGCTCGCGCGCTGAAG 386  
QY 302 AlaPheSerGlyLeuAlaHisArgCysGlnTrpValArgGluArgGlnGlyValSerTyr 321  
Db 385 GCGTTTTCGCGCTGCTCATCGCTGCCAGTGGGTACGCGAGCGCGAGCGCTGAGCTAC 326  
QY 322 TyrAspAspSerLysAlaThrAsnValGlyAlaAlaLeuAlaAlaLeuGlyLeuGly 341  
Db 325 TACGACGATTCCAAGGCCACCAACGTCGCGCGCGCGCTGGCGCGATCGAGGGGCTGGGT 266  
QY 342 AlaAspIleAspGlyLysLeuValLeuAlaGlyLysArgGlyAspGlyLysGlyAlaAspPhe 361  
Db 265 GCCGACATCGACGCAAGCTGGTGTCTCGCGCGGAGAGCGCAAGGCGCGCGATTTC 206  
QY 362 HisAspLeuArgGluProValAlaAlaArgPheCysArgAlaValValLeuLeuGlyArgAsp 381  
Db 205 CATGACCTCGCGAGCGCGTCTGCTGCCGCGCGCTTCTGCCGCGCGTGTACTGCTGGCGCTGAC 146  
QY 382 AlaGlyLeuIleAlaGlnAlaLeuGlyAsnAlaValProLeuValArgValAlaThrLeu 401  
Db 145 GCGGGCTGATGGCCAGCAGCTGGCAACGCGGTACCGCTGGTGGCGCTGCCAACGCTG 86  
QY 402 AspGluAlaValArgGlnAlaAlaGluLeuAlaArgGluGlyAspAlaValLeuLeuSer 421  
Db 85 GACGAGCAGCTCCGCGAGCGCGCGAGCTGGCGCGCGCGCGCGCGCGCGCGCGCGCTG 26  
QY 422 ProAlaCysAlaSerLeuAspMet 429  
Db 25 CCGGCTGCGCGAGCGCTGGACATG 2

RESULT 5  
US-09-557-884-1  
; Sequence 1, Application US/09557884  
; Patent No. 6506581  
; GENERAL INFORMATION:  
; APPLICANT: Fleischmann et al.  
; TITLE OF INVENTION: The Nucleotide sequence of the Haemophilus influenzae Rd Genome, Fragments Thereof, and Uses Thereof  
; NUMBER OF SEQUENCES: 1  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Human Genome Sciences, Inc.  
; STREET: 9410 Key West Avenue  
; CITY: Rockville  
; STATE: MD  
; COUNTRY: USA  
; ZIP: 20850  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 3 1/2 inch diskette  
; COMPUTER: Dell Pentium  
; OPERATING SYSTEM: MS DOS v6.22  
; SOFTWARE: ASCII Text  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/557,884  
; FILING DATE: 25-Apr-2000  
; CLASSIFICATION: <Unknown>  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/476,102  
; FILING DATE: JUN-5-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Michelle S. Marks  
; REGISTRATION NUMBER: 41,971  
; REFERENCE/DOCKET NUMBER: PB186P3  
; TELECOMMUNICATION INFORMATION:



Db	1203810	CATACCGTT-----TTTTTTCGGGAAAAATAGTCGGAT-----TATTTGGCTA	1203851
Qy	247	IleclucIuAspGlyGlnLysTrpLeuAlaPheGlnPheAspLysLeuLeuProValGly	266
Db	1203852	AAACCTGAATGGCAGCACTAATTTAATGGTTAAAGATGATGATTTTACCTGTGAA	1203911
Qy	267	GlulLysIleArgGlyAlaHisAsnTySerAsnAlaLeuAlaLeuAlaLeuGly	286
Db	1203912	GAAGCTACATGCTTGGTCGCCATAATATATGAACATTTTGGCAGCAACAGCATTTGGCA	1203971
Qy	287	HisAlaValcGlyLeuProPheAspAlaMetLeuGlyAlaLeuLysAlaPheSerGlyLeu	306
Db	1203972	CACGCTATAGGTATTAAATTTAGATTCAATTCGTACCGCACTTCGCTATTTCAAAGGGTTA	1204031
Qy	307	AlaHisArgCysGlnTrpValArgGluArgGlnGlyValSerTyrTyrAspAspSerLys	326

DQ	1204032	GATCATCGTTTTCAATTAGTGGCATCAAGCTAATTTGGCATTTCGTTGGATTAAATGAAGACTCTAAA		1204099
QY	327	AlaThrAsnValGlyAlaAlaLeuAlaAlaIleGluGlyLeuGlyAlaAspIleAspGly	346	
		::::::::::::::::::::		
Db	1204092	GCAACAAATGTGGGGAGTACAGTTGTGTCATTGCCTGGCGCTT-----TATATTGAGGGT	1204145	
QY	347	LysLeuValLeuLeuAlaGlyGlyAspGlyLysGlyAlaAspPheHisAspLeuAsgLys	366	
		::::::::::::::::::::		

DD 1204146 AATATGCAATTGTTGCTAGGCGGAGACGGAAAGGGGCTGATTTTTCAGAAATTAGCTGAA 1204203

Qy	367	PROVALAATG	:: :: :: :: :: :: :: :: :: ::	PNECYSGALVAIVALLEULEUGIYALGASP	1204250
Dy	1204206	TAAATTAATCAACACACATTTATTGTTATGT	-----	TTTGGTCGAGAT	1204250

QY 30Z ATAGYU LEUTTEATAGIMVATALEUGYASNAITAVATPOTLEUVATIGVATIAHILEU 401  
|||:::||:::  
::

DD IZ04ZJ1 GGIGCCICICICACAAIIIIICACCAAGIIAI-----11G IZ04Z09

QY 40Z ASPGUAIVBIAFGGINAIBAGIULEUAIABG-----GIUGLYASPAIA 41/

DB 1204290 TTCGATACAAATGGGAACAAGCGATAGAAATTCTTACGCCCAACAATTCGCAAAAGCGGAGATATG 1204349

```
0Y      valbeusefPloAtAcYsAtaseIbeuAsrmeCneLcYsAsnHeGluGlUArgLy 437
```

DD IZ04330 GIATTAI6IC9CCIGCIIGTGCAGGTCICGATCAGTITGCTTCTTTGGAAAGGCCCGCGC IZ04400

αγ 430 αγνευσε 440  
|||

DD 1204410 040504111 1204410

RESOLUTION /  
US-09-252-991A-7928/c

; sequence 1928, application US/09232991A  
; Patent No. 6551795

; APPLICANT: Marc J. Rubenfield et al.

**; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS**

; FILE REFERENCE: 10/130.130  
; CURRENT APPLICATION NUMBER: US/09/252,991A

7 CURRENT FILING DATE: 1999 04 10  
; PRIOR APPLICATION NUMBER: US 60/074,788

; PRIOR APPLICATION NUMBER: US 60/094,190

NUMBER OF SEQ ID NOS: 33142

; LENGTH: 564

ORGANISM: *Pseudomonas aeruginosa*

[illegible]

Pred. No.:	2.01e-60	Length:	564
------------	----------	---------	-----

Percent Similarity: 100.00%

Conservative: 0



Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 27.70% Indels: 0  
DB: 4 Gaps: 0

US-09-701-229-2 (1-448) x US-09-252-991A-7928 (1-564)

```
QY 322 TyrAspSerLysAlaThrAsnValGlyAlaAlaLeuAlaAlaGluGlyLeuGly 341
D 563 TACGACGATTCCAAAGGCCACCAACGTCGGCCGCCCTGGCGGCGATCGAGGGCTGGGT 504
QY 342 AlaAspLeuAspGlyLysLeuValLeuAlaGlyAlaGlyLysGlyAlaAspPhe 361
D 503 GCCGACATCGACGCGAGCTGGTGTCTGCCCGCGGACGCGCAAGGGCGCGATTC 444
QY 362 HisAspLeuArgGluProValAlaAlaArgPheCysArgAlaValLeuLeuGlyArgAsp 381
D 443 CATGACCTGCGGAGCGGTCGCGGCTTCTGCCGGGGGGTACTGCTTGGCGCGTGAC 384
QY 392 AlaGlyLeuLeAlaGlnAlaLeuGlyAsnAlaValProLeuValArgValAlaThrLeu 401
D 383 GCCGGGCTGATTGCCAGGCACCTGGCAACGCGGTACCGCTGTGGCGCTGCCAACGGCTG 324
QY 402 AspGluAlaValArgGlnAlaAlaGluLeuAlaArgGluGlyAspAlaValLeuLeuSer 421
D 323 GACGAAGCAGTCCCGCAGCGCGGCTGCTGCCCGCGGAGCGATGCGGCTGTGTTCG 264
QY 422 ProAlaCysAlaSerLeuAspMetPheLysAsnPheGluGluArgGlyArgLeuPheAla 441
D 263 CCGGCTGCGGAGCGCTGGACATGTTCAAGAACTTCGAGAACCGGACGCGCTGTTCGCC 204
QY 442 LysAlaValGluGluLeuAla 448
D 203 AAAGCCGTAGGAGCTAGCG 183
```

RESULT 8

US-08-934-481-1  
; Sequence 1, Application US/08934481  
; Patent No. 5929045  
; GENERAL INFORMATION:  
; APPLICANT: Wallis, Nichola G.  
; APPLICANT: Fueyo, Joanna L.  
; APPLICANT: Lonetto, Michael A.  
; TITLE OF INVENTION: NOVEL MURD  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Dechert Price & Rhoads  
; STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre  
; CITY: Philadelphia  
; STATE: PA  
; COUNTRY: US  
; ZIP: 19103  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/934,481  
; FILING DATE:  
; CLASSIFICATION: 514  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Dickinson, Todd O  
; REGISTRATION NUMBER: 28,354  
; REFERENCE/DOCKET NUMBER: GM10070A  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 215-994-2252  
; TELEFAX: 215-994-2222  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:

; LENGTH: 2193 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
US-08-934-481-1  
Alignment Scores:  
Pred. No.: 5,08e-57 Length: 2193  
Score: 603.50 Matches: 164  
Percent Similarity: 53.16% Conservative: 80  
Best Local Similarity: 35.73% Mismatches: 178  
Query Match: 26.70% Indels: 37  
DB: 2 Gaps: 13

US-09-701-229-2 (1-448) x US-08-934-481-1 (1-2193)

```
QY 7 AspHisPheArg-----IleValValGlyLeuGlyLysSerGlyMetSerLeu 22
D 494 GATCAATTTAAATAATAAGAAAGTCTTGTGTTAGGTTGGCCAGTCTGGTGAATCTGCA 553
QY 23 ValArgTyrLeuAlaAlaArgArgGlyLeuProPheAlaValAlaAspThrArg----- 39
D 554 GCTCGTTTGTGGACAAGCTAGTGCCATTTGTGACAGTAATATGATGGGAACCTTTCGAG 613
QY 40 GluAsnProProGluLeuAlaThrLeuArgAlaGlnTyrProGlnValGluValArgCys 59
D 614 GACAATCCAGCTGCCCAAAGTTTGTCTGGAA-----GAAGGATCAAGTCAATTACA 664
QY 60 GlyGlu-----LeuAspAlaGluPheLeuCysSerAlaArgGluLeuTyr 74
D 665 GGTGCCCATCTCTTGGAACTCTTGATGAAGAGTTTGTCTTATGGTGA----- 715
QY 75 ValSerProGlyLeuSerLeuArgThrProAlaLeuValGlnAlaAlaLysGlyVal 94
D 716 ---AATCCAGTATCCCTACAAATCCCATGATGAAAGGCTTTCGCAAGGGAAT 772
QY 95 ArgIleSerGlyAspIleAspLeuPheAlaArgGluAlaLysAlaProIleValAlaLe 114
D 773 CCAGTCTTGACTGAGTGGATTTGCTTATTTGATTTCAGAACGACCATGATTGTTATC 832
QY 115 ThrGlySerAsnAlaLysSerThrValThrThrLeuValGlyGluMetAlaValAlaAla 134
D 833 ACAGGATCGAAGCGTAAAGAACCAACACGACTATGATGGGGAAGTTTGTACTGCTCT 892
QY 135 AspLysArgValAlaValGlyGlyAsnLeuGlyThrProAlaLeuAspLeu-----Leu 152
D 893 GGGCAACATGCTCTTTATCAGGGAATATCGGCTATCCAGCTAGTCAAGTGGCCAAACT 952
QY 153 AlaAspAspIleGluLeuTyrValLeuGluLeuSerSerPheGlnLeuGluThrCysAsp 172
D 953 CGCTCGGCAAGGACACGCTTGTATGGAACCTTCTCTTCCAACTCATGACTGACCAT 1012
QY 173 ArgLeuAsnAlaGluValAlaThrValLeuAsnValSerGluAspHisMetAspArgTyr 192
D 1013 GAATTTTCATCCAGAGATTGCGGTTATTACCAACCTTATGCCAACTCATATGACTACCAT 1072
QY 193 AspGlyMetAlaAspTyrHisLeuAlaLysHisArgIle-----PheArgGlyAla 209
D 1073 GGGTCATTTTCGGAATATGTAGCAGCCCAAGTGAATATCCAGAACAGATGACACAGCT 1132
QY 210 ArgGlnValValAlaAsnArgAlaAspAlaLeuThrArgProLeuIleAlaAspThrVal 229
D 1133 GATTTCTCTTGTGTAAGCTTTAATCAAGACTTTGACTTCCAAAGACAGACCACTGTTGTA 1192
QY 230 ProCysTrpSerPheGlyLeuAsnLysProAspPheLysAlaPheGlyLeuIleGluGlu 249
D 1193 CCATTTTCAACA-----CTTGAAAGAGTTGAT-----GGAGCTTATCTCGAA 1234
QY 250 AspGlyGlnLysTrpLeuAlaPheGlnPheAspLysLeuLeuProValGlyGluLys 269
D 1235 GATGGTCAA-----CTCTACTTCGGTGTGAAGTAGTATGTCGGCAGCAAGTAATCGGT 1288
QY 270 IleArgGlyAlaHisAsnTyrSerAsnAlaLeuAlaLeuAlaLeuGlyHisAlaVal 289
```

```
Db 1289 GTTCCAGGTAGCCACCAATGGAATGCGCTTGCAGCTATTGCTGTAGCCAAAGCTTGGT 1348
QY 290 GlyLeuProPheAspAlaMetLeuGlyAlaLeuLysAlaPheSerGlyLeuAlaHisArg 309
Db 1349 GGTGTGGACAAATCAACCAATCAAGGAACCTTTTCAGCCTTCGGTGTGCAACACCGT 1408
QY 310 CysGlnTrpValArgGluArgGlnGlyValSerTyrTyrAspSerLysAlaThrAsn 329
Db 1409 CTCAGTTGTGGATGACATCAAGGGTGTAAATTTCTATAACGACAGTAATCAACTAAT 1468
QY 330 ValGlyAlaAlaLeuAlaLeuGlyAlaLeuGlyAlaAspLeuAspGlyLysLeuVal 349
Db 1469 ATCTTGCTACTCAAAAGCCCTTGTGAGGATTGAC-----AACAGCAGGTGCTGTC 1519
QY 350 LeuLeuAlaGlyGlyAspGlyLysGlyAlaAspPheHisAspLeuArgGluProValAla 369
Db 1520 TTGATGTCAGGTGGTTGGACCGTGCATGAGTTTGACAAATG---GTGCCAGACAT 1576
QY 370 ArgPheCysArgAlaValValLeuLeuGlyArgAspAlaGlyLeuIleAlaGlnAlaLeu 389
Db 1577 ACTGGACTCAAGAAGATGTCATCTGCTGCAAGCTGTCAAAACGGCGAGCA 1636
QY 390 GlyAsnAla---ValProLeuValArgValAlaAlaThrLeuAspGluAlaValArgGlnAla 408
Db 1637 GACAAGCCTGGTGTCTGCTTATGTGGAGCGACAGATATTGCGAGTCCGACCGCAAGGCC 1696
QY 409 AlaGluLeuAlaArgGluGlyAspAlaValLeuLeuSerProAlaCysAlaSerLeuAsp 428
Db 1697 TATGACCTGGCAGTCAAGAGATGTTGCTTCTTAGTCCCTGCCAATGCCAGCTGGAT 1756
QY 429 MetPheLysAspPheGluGluArgGlyArgLeuPheAlaLysAlaValGluLeu 447
Db 1757 ATGTATGCTAACTTTCAAGTACGTGGCAGCTCTTTATCGACACAGTAGCGGAGTTA 1813
```

## RESULT 9

```
US-09-290-602-1
; Sequence 1, Application US/09290602
; Patent No. 6350598
; GENERAL INFORMATION:
; APPLICANT: Wallis, Nicholas G.
; APPLICANT: Fuenyo, Joanna L.
; APPLICANT: Lonetto, Michael A.
; TITLE OF INVENTION: NOVEL MURD
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dechert Price & Rhoads
; STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre
; CITY: Philadelphia
; STATE: PA
; COUNTRY: US
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/290,602
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/934,481
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Dickinson, Todd Q.
; REGISTRATION NUMBER: 28,354
; REFERENCE/DOCKET NUMBER: GM10070A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-994-2252
; TELEFAX: 215-994-2222
; TELEX:
; INFORMATION FOR SEQ ID NO: 1:
```

SEQUENCE CHARACTERISTICS:  
LENGTH: 2193 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
US-09-290-602-1

Alignment Scores:  
Pred. No.: 5,08e-57 Length: 2193  
Score: 603.50 Matches: 164  
Percent Similarity: 53.16% Conservative: 80  
Best Local Similarity: 35.73% Mismatches: 178  
Query Match: 26.70% Indels: 37  
DB: 4 Gaps: 13

US-09-701-229-2 (1-448) x US-09-290-602-1 (1-2193)

QY 7 AspHisPheArg-----IleValValGlyLeuGlyLysSerGlyMetSerLeu 22  
Db 494 GATCAATTTAAATAAAGAAAGTCTTGTAGTTTGGCCAAAGTCTGCTGAATCTGCA 553  
QY 23 ValArgTyrLeuAlaArgArgGlyLeuProPheAlaValAlaAspThrArg----- 39  
Db 554 GCTCGTTTGTGGACAGCTAGTGCCATGTGACAGTAATGATGGGAACCTTTCGAG 613  
QY 40 GluAsnProProGluLeuAlaThrLeuArgAlaGlnTyrProGlnValGluValArgCys 59  
Db 614 GACAATCCAGCTGCCCAAAGTTTGTCTGGAA-----GAAGGATCAAGGTCAATTACA 664  
QY 60 GlyGlu-----LeuAspAlaGluPheLeuCysSerAlaArgGluLeuTyr 74  
Db 665 GTGGCCATCTTGGAACTCTTGGATGAAGATTTTGTCTTATGTGTGAA----- 715  
QY 75 ValSerProGlyLeuSerLeuArgThrProAlaLeuValGlnAlaAlaLysGlyVal 94  
Db 716 ---ATTCAGGTATCCCTTACAACATCCATGATTGAAAGGCTTTGCCAAGGAAT 772  
QY 95 ArgIleSerGlyAspIleAspLeuPheAlaArgGluAlaLysAlaProIleValAlaIle 114  
Db 773 CCAGTCTTCTGCTGAGGTGGAATTTGGCTTTATTTGATTTCAGAAGCACCGATTATGTTATC 832  
QY 115 ThrGlySerAsnAlaLysSerThrValThrLeuValGlyGluMetAlaValAlaAla 134  
Db 833 ACAGGATCGAACGGTAAGAACACACACACTGATGATTGGGGAAGTTTGTACTGCTGCT 892  
QY 135 AspLysArgValAlaValGlyGlyAsnLeuGlyThrProAlaLeuAspLeu-----Leu 152  
Db 893 GGGCAACATGCTTTTATCAGGGAATATCGGCTATCCAGCTAGTCAAGTGGCCCAACT 952  
QY 153 AlaAspAspIleGluLeuTyrValLeuGluLeuSerSerPheGlnLeuGluThrCysAsp 172  
Db 953 GCGTCGGACAAAGCACACGCTTCTTATGGAACCTTCTCTTCCCACTCATGGTGTTCAA 1012  
QY 173 ArgLeuAsnAlaGluValAlaThrValLeuAsnValSerGluAspHisMetAspArgTyr 192  
Db 1013 GAATTTTCATCAGAGATTGCGGTATTACCAACCTTATGCCAACCCTATATCGACTACCAT 1072  
QY 193 AspGlyMetAlaAspTyrHisLeuAlaLysHisArgIle-----PheArgGlyAla 209  
Db 1073 GGGTCATTTTCCGAATATGTACACCAAGTGGATATATCCAGAACAGATGACAGCAGCT 1132  
QY 210 ArgGlnValValAlaAsnArgAlaAspAlaLeuThrArgProLeuIleAlaAspThrVal 229  
Db 1133 GATTTTCTTGTCTTGAACCTTTAATCAAGACTTTGACTTCCAGACAGAACCCACTGTTGTA 1192  
QY 230 ProCysTrpSerPheGlyLeuAsnLysProAspPheLysAlaPheGlyLeuIleGluGlu 249  
Db 1193 CCATTTTCAACA-----CTTGAAGAGGTGAT-----GGAGCTTATCTGGAA 1234  
QY 250 AspGlyGlnLysTrpLeuAlaPheGlnPheAspLysLeuLeuProValGlyGluLeuLys 269  
Db 1235 GATGTCAA-----CTCTACTTCTGGTGTGAAGTAGTATCGGACGCAAGTAATCGGT 1288

```

Qy 270 IleArgGlyAlaHisAsnTyrSerAsnAlaLeuAlaLeuAlaLeuGlyHisAlaVal 289
Db   : |||:||||| |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
1289 GTTCAGGACCCACAAATGGAATGCGCTATGCGGACTATGCTGTAGCAACGCTCGT 1348
Qy 290 GlyLeuProPheAspAlaMetLeuGlyAlaLeuLysAlaPheSerGlyLeuAlaHisArg 309
Db   : |||:||||| |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
1349 GGTGTGGACAATCAAAACCAATCAAGGAAACTCTTTTCAGCCCTCGGTGTGTCACACCGT 1408
Qy 310 CysGlnTrpValArgGluAlaGlnGlyValSerTyrTyrAspSerLysAlaThrAsn 329
Db   : |||:||||| |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
1409 CTCAGTTTGGGTGATGACATCAAGGGTGTAAATCTATACGACATTAATCACTAAT 1468
Qy 330 ValGlyAlaLeuAlaAlaIleGluGlyLeuGlyAlaAspLeuLysLeuVal 349
Db   : |||:||||| |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
1469 ATCTTGGCTACTCAAAAGCCTGTGACGATTTGAC-----AACAGCAAGTGTGTC 1519
Qy 350 LeuLeuAlaGlyAspGlyLysGlyAlaAspPheHisAspLeuArgGluProValAla 369
Db   : |||:||||| |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
1520 TTGATTCAGGTGTTGGACCGTGGCAATGAGTTTGACGAATTG---GTGCCAGACATT 1576
Qy 370 ArgPheCysArgAlaValValLeuLeuGlyArgAspAlaGlyLeuIleAlaGlnAlaLeu 389
Db   : |||:||||| |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
1577 ACTGGACTCAAGAAGATGGTCATCTGGTGGTCAATCTGCACAGCTGTCAACGGCGACGA 1636
Qy 390 GlyAsnAla---ValProLeuValArgValAlaThrLeuAspGluAlaValArgGlnAla 408
Db   : |||:||||| |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
1637 GACAAGGCTGTGTGCTGATGTGGAGCGCACAGATATGCAGATGGCGACCGCAAGGCC 1696
Qy 409 AlaGluLeuAlaArgGluGlyAspAlaValLeuLeuSerProAlaCysAlaSerLeuAsp 428
Db   : |||:||||| |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
1697 TAGAGCTTCGGACTCAAGGAGATGTGTTCTTCTTAGTCTGCCAATGCCAGTGGGAT 1756
Qy 429 MetPheLysAsnPheGluGluArgGlyArgLeuPheAlaLysAlaValGluLeu 447
Db   : |||:||||| |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
1757 ATGTATCTAACTTTGAAGTACGTGGCGACCTCTTTATCGACACAGTAGCGGAGTTA 1813

```

# RESULT 10

```

US-08-961-527-195/c
; Sequence 195, Application US/08961527
; Patent No. 6420135
; GENERAL INFORMATION:
; APPLICANT: Charles Kunsch
; TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences
; NUMBER OF SEQUENCES: 391
; CORRESPONDENCE ADDRESS:
; ADDRESS: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/961,527
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 195:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6693 base pairs

```

```

; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
US-08-961-527-195

Alignment Scores:
Pred. No.: 2.81e-55 Length: 6693
Score: 594.50 Matches: 163
Percent Similarity: 53.13% Conservative: 83
Best Local Similarity: 35.21% Mismatches: 176
Query Match: 26.31% Indels: 41
DB: 4 Gaps: 14

US-09-701-229-2 (1-448) x US-08-961-527-195 (1-6693)

Qy 7 AsphHisPheArg-----lIleValValGlyLeuGlyLysSerGlyMetSerLeu 22
Db   : |||:||||| |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
3784 GATCAATTTAAATAAAGAGTCTCTTTAGTTTGGCCAAAGTCTGGTGAATCTGCA 3725
Qy 23 ValArgTyrLeuAlaArgArgGlyLeuProPheAlaValAlaAspThrArg----- 39
Db   : |||:||||| |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
3724 GCTGTTTGTGGACAGCTAGGTGCCATTCTGACAGTAATGATGGAAACCTTTCGAG 3665
Qy 40 GluAsnProProGluLeuAlaThrLeuArgAlaGlnTyrProGlnValGluValArgCys 59
Db   : |||:||||| |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
3664 GACAATCCAGCTGCCAAAGTTTCTGTGAA-----GAAGGATCAAGTCAATACA 3614
Qy 60 GlyGlu-----LeuAspAlaGluPheLeuCysSerAlaArgGluLeuTyr 74
Db   : |||:||||| |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
3613 GGTGGCCATCCTTTGGAACCTCTTGGATGAAGAGTTTGCCCTTATGGTGAA----- 3563
Qy 75 ValSerProGlyLeuSerLeuArgThrProAlaLeuValGlnAlaAlaLysGlyVal 94
Db   : |||:||||| |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
3562 ---AATCCAGTATCCCTACAAATCCCATGATTGAAAGGCTTTGGCCAAAGGAATT 3506
Qy 95 ArgIleSerGlyAspIleAspLeuPheAlaArgGluAlaLysAlaProIleValAlaIle 114
Db   : |||:||||| |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
3505 CCAGTCTTGACTGAGGTGGAATGCTTATTGATTTCAGAAAGCAGCATTTATTGGTATC 3446
Qy 115 ThrGlySerAsnAlaLysSerThrValThrThrLeuValGlyGluMetAlaValAlaAla 134
Db   : |||:||||| |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
3445 ACAGATCGAAGGTGAGAACACCAACGACTATGATGGGGAAGTTTTCAGTCTGCT 3386
Qy 135 AspLysArgValAlaValGlyGlyAsnLeuGlyThrProAlaLeuAspLeu-----Leu 152
Db   : |||:||||| |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
3385 GGCCAACATGCTCTTTTATCAGGCAATATCGCTATCCAGCTAGTCAGTTGCTCAATA 3326
Qy 153 AlaAspAspIleGluLeuTyrValLeuGluLeuSerSerPheGlnLeuGluThrCysAsp 172
Db   : |||:||||| |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
3325 GCATCAGATAAGGACACGCTTGTGTAAGCTTTCTTTTCCAACTCATCGGTGTTCAA 3266
Qy 173 ArgLeuAsnAlaGluValAlaThrValLeuAsnValSerGluAspHisMetAspArgTyr 192
Db   : |||:||||| |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
3265 GAATTCATCCAGAGATTCGGTTATTACCACCTCATGCACTATCATCATCATCATCAT 3206
Qy 193 AspGlyMetAlaAspTyrHisLeuAlaLysHisArgIle-----PheArgGlyAla 209
Db   : |||:||||| |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
3205 GGGTCATTTTCGGAATATGATGACGCCCAAGTGAATATCCAGAACAAAGATCACAGCAGCT 3146
Qy 210 ArgGlnValValValAsnArgAlaAspAlaLeuThrArgProLeuIleAlaAspThr--- 228
Db   : |||:||||| |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
3145 GATTTCCTTGTCTGAAGCTTTAATCAAGACTTGGCAAAAGACTTTCAGTTCACAGACGAA 3086
Qy 229 -----ValProCysTyrSerPheGlyLeuAsnLysProAspPheLysAlaPheGly 245
Db   : |||:||||| |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
3085 GCCACTCTGTACCATTTTCAACA-----CTTGAAAGGTTGAT-----GGA 3044
Qy 246 LeuIleGluGluAspGlyGlnLysTrpLeuAlaPheGlnPheAspLysLeuLeuProVal 265
Db   : |||:||||| |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
3043 GCTTATCTGGAAGATGTCAA-----CTTACTCTCGTGTGAGTAGTATGATGCGACGC 2990
Qy 266 GlyGluLeuLysIleArgGlyAlaHisAsnTyrSerAsnAlaLeuAlaLeuAlaLeu 285
Db   : |||:||||| |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||

```

```
Db 2989 AATGAATCGGTGTTCCAGTACCCACCAATGTGAAATGCCCTTCGGACTATTGCTGTA 2930
QY 286 GlyHisAlaValGlyLeuProPheAspAlaMetLeuGlyAlaLeuLysAlaPheSerGly 305
Db 2929 GCCAAGCTTCGTGATGGGACAAATCAACCAATCAAGGAAACTCTTTCAGCCCTTCGGTGGT 2870
QY 306 LeuAlaHisArgCysGlnTrpValArgGluArgGlnGlyValSerTyrTyrAspAspSer 325
Db 2869 GTCAAAACACCGCTCCAGTTGTGGATGACATCAAGGGTGTAAATCTATACGACAGT 2810
QY 326 LysAlaThrAsnValGlyAlaAlaLeuAlaAlaIleGluGlyLeuGlyAlaAspIleAsp 345
Db 2809 AATCAACTATATCTGGCTACTCAAAAGCCTTGTGAGGATTTGAC-----AAC 2759
QY 346 GlyLysLeuValLeuLeuAlaGlyGlyAspGlyLysGlyAlaAspPheHisAspLeuArg 365
Db 2758 AGCAAGTCGTCTGTGATTTGAGGTGCTTGGACCGTGGCAATGAGTTTGACGAATG--- 2702
QY 366 GluProValAlaArgPheCysArgAlaValValLeuLeuGlyArgAspAlaGlyLeuIle 385
Db 2701 GTGCCAGACATTACTGACTCAAGAAAGATGGTCACTCGGTCAATCTGCAGAACGTGTC 2642
QY 386 AlaGlnAlaLeuGlyAsnAla---ValProLeuValArgValAlaThrLeuAspGluAla 404
Db 2641 AAACGGGACGACGACAAAGCTGTGCTTATGTGAGCGGACACATATTCAGATGCG 2582
QY 405 ValArgGlnAlaAlaGluLeuAlaArgGluGlyAspAlaValLeuLeuSerProAlaCys 424
Db 2581 ACCCGCAAGCCTATCAGCTTGGACTCAAGGAGATGTGTTCTTCTAGTCTGCAAT 2522
QY 425 AlaSerLeuAspMetPheLysAsnPheGluGluArgGlyArgLeuPheAlaValAla 444
Db 2521 GCTAGCTGGGATATGATCTAACTTTGAATACGTGGGACCTCTTTATCGACACAGTA 2462
QY 445 GluGluLeu 447
Db 2461 GCGGAGTTA 2453

RESULT 11
US-09-252-991A-7552
; Sequence 7552, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR FILING DATE: 1998-02-18
; PRIOR FILING DATE: 1998-02-18
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 7552
; LENGTH: 360
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-7552

Alignment Scores:
Pred. No.: 7,73e-57 Length: 360
Score: 591.00 Matches: 119
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 26.15% Indels: 0
DB: 4 Gaps: 0

US-09-701-229-2 (1-448) x US-09-252-991A-7552 (1-360)

QY 5 AlaSerAspHisPheArgIleValValGlyLeuGlyLysSerGlyMetSerLeuValArg 24
Db 3 GCCTCCGACCACTTCGGATCGTTGTCGCCCTCGCGAAGAGCGGCATGTCCCTGCTGCGC 62
```

```
QY 25 TyrLeuAlaArgArgGlyLeuProPheAlaValAlaAspThrArgGluAsnProProGlu 44
Db 63 TACCTGGCGCGCGCGCTTGCCTTTCGCCGTTGTGCATACCCGAGAGAACCCGCCGGAG 122
QY 45 LeuAlaThrLeuArgAlaGlnTyrProGlnValGluValArgCysGlyGluLeuAspAla 64
Db 123 CTGGCCACCGCTGCGTCCCGAGTATCGCAGTATCGCAGTGGAAAGTGGTTCGCGCGAACCTCGACGCC 182
QY 65 GluPheLeuCysSerAlaArgGluLeuTyrValSerProGlyLeuSerLeuArgThrPro 84
Db 183 GAGTTCTCTGCTCCCGCCCGCAACTCTAGCTCAGCCCCCGCTGTGCTGCGCACCACCCC 242
QY 85 AlaLeuValGlnAlaAlaAlaLysGlyValArgIleSerGlyAspIleAspLeuPheAla 104
Db 243 GCCTGTTACAGCGCGCGCGAAGGTGTGCGCATCTCCGGTGCATCTCGATCTCTTCGCC 302
QY 105 ArgGluAlaLysAlaProIleValAlaIleThrGlySerAsnAlaLysSerThrVal 123
Db 303 CGCGAGCGAAGCGCGCGATCGTGCATCACCGGTTCCAAACGCGAAGAGCACCGTG 359
```

## RESULT 12

```
US-08-665-435A-1
; Sequence 1, Application US/08665435A
; Patent No. 5681694
; GENERAL INFORMATION:
; APPLICANT: Skatrud, Paul
; APPLICANT: Peery, Robert
; APPLICANT: Hoskins, JoAnn
; APPLICANT: Wu, Chyun-yeh Earnest
; TITLE OF INVENTION: Biosynthetic Gene MurD of Streptococcus
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Eli Lilly and Company
; STREET: Lilly Corporate Center
; CITY: Indianapolis
; STATE: Indiana
; COUNTRY: US
; ZIP: 46285
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/665,435A
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Webster, Thomas D
; REGISTRATION NUMBER: 39,872
; REFERENCE/DOCKET NUMBER: X-9900
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 317-276-3334
; TELEFAX: 317-276-3861
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1350 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1350
US-08-665-435A-1
```

```
Alignment Scores:
Pred. No.: 1.86e-55 Length: 1350
Score: 596.50 Matches: 162
```

Percent Similarity:	52.92%	Conservative:	83
Best Local Similarity:	34.99%	Mismatches:	177
Query Match:	25.95%	Indels:	41
DB:	1	Gaps:	14

US-09-701-229-2 (1-448) x US-08-665-435A-1 (1-1350)

Qy	7	AspHisPheArg-----IleValValcGlyLeuGlyLysSerGlyMetSerLeu	22
Db	13	GATCAATTTTTAAAAATAAGAAAGTCTTGTGTTTAGTTGGCCCAAGTCTGGTGAATCTGCA	72
Qy	23	ValArgTyrLeuAlaAaArgAGGlyLeuProPheAlaValAlaAspThrArg-----	39
Db	73	GCTCGTTGTTGGACAAGCTAGGTGCCATTCTGCACAGTAAATGATGGGAACCTTTCGAG	132
Qy	40	GluaSnProGluLeuAlaThrLeuArgAlaGlnTyrProGlnValGluValArgCys	59
Db	133	GACAATCCAGCTGCCCAAGGTTTGCTCGAA-----GAAGGGATCAAGGTCAATTACA	183
Qy	60	GlyGlu-----LeuAspAlaGluPheLeuCysSerAlaArgGluLeuTyr	74
Db	184	GTGGGCCATCCTTTGGAACTCTGGATGAAGAGTTTGCCCTATATGGTGAAA-----	234
Qy	75	ValSerProGlyLeuSerLeuArgThrProAlaLeuValGlnAlaAlaLysGlyVal	94
Db	235	---AATCCAGGTATCCCTCTCAACAATCCCATGATGATGAAAGGGTTTGCCCAAGAGAAT	291
Qy	95	ArgTleSerGlyAspIleAspLeuPheAlaArgGluAlaLysAlaProIleValAlaIle	114
Db	292	CCAGCTGTGACTGAGGTGGAAATGGCTTATTTGATTTCAGAAGCAACCGATTATTGGTATC	351
Qy	115	ThrGlySerAsnAlaLysSerThrValThrLeuValGlyGluMetAlaValAlaAla	134
Db	352	ACAGGATCGAACGGTAAGACAAACACACGACTATGATTTGGGAAGTTTGTACTGCTGCT	411
Qy	135	AspLysArgValAlaValAlGlyGlyAsnLeuGlyThrProAlaLeuAspLeu-----	152
Db	412	GGGCAACATGGTCTTTTATCAGGGAATATCGGCTATCTCGCAGTCAGGTGCTCAAAATA	471
Qy	153	AlaAspAspIleGluLeuTyrValLeuGluLeuSerSerPheGlnLeuGluThrCysAsp	172
Db	472	GCATCAGATAAGGACACGCTGTGTTATGGAACCTTTCTCTTCCAACTCATGGGTGTTCAA	531
Qy	173	ArgLeuAsnAlaGluValAlaThrValLeuAsnValSerGluAspHisMetAspArgTyr	192
Db	532	GAATTCATCCAGAGATGCGGTATTATTACCAACCTCATCGCAACTCATATCGACTACC	591
Qy	193	AspGlyMetAlaAspTyrHisLeuAlaLysHisArgIle-----PheArgGlyAla	209
Db	592	GGGTCAITTTCTGAATATGTAGCAGCCCAAGTGGATATCCAGAACAGATGACAGCGT	651
Qy	210	ArgGlnValValValAsnArgAlaAspAlaLeuThrArgProLeuIleAlaAspThr---	228
Db	652	GATTTCTTTGTTCTTGAACCTTTAATCAAGACTTGGCAAAAGACTTGACTTCCAAAGACA	711
Qy	229	-----ValProCysTyrSerPheGlyLeuAsnLysProAspPheLysAlaPheGly	245
Db	712	GCCACTGTTGTACCATTTCACAA-----CTTGAAAAGTTGAT-----GGA	753
Qy	246	LeuIleGluAspGlyGlnLysTyrLeuAlaPheGlnPheAspLysLeuLeuProVal	265
Db	754	GCTTATCTGGAAGATGGTCAA-----CTCTACTTCCGTGGTGAAGTAGTCAATGCAGCG	807
Qy	266	GlyGluLeuLysIleArgGlyAlaHisAsnTyrSerAsnAlaLeuAlaAlaLeuAlaLeu	285
Db	808	AATGAATCCGTTGTTCCAGGTAGCCACAACTTGGAAAATGCCCTTTCGCGACTATTGCTGT	867
Qy	286	GlyHisAlaValGlyLeuProPheAspAlaMetLeuGlyAlaLeuLysAlaPheSerGly	305
Db	868	CCCAAGCTTCGTGATGTGGACAAATCAACCATCAAGGAAACTCTTTTCAGCCCTTCGGTGT	927
Qy	306	LeuAlaHisArgCysGlnTrpValArgGluArgGlnGlyValSerTyrTyrAspAspSer	325

RESULT 13

```

US-08-665-435A-3
; Sequence 3, Application US/08665435A
; Patent No. 5681694
; GENERAL INFORMATION:
; APPLICANT: Skatrud, Paul
; APPLICANT: Peery, Robert
; APPLICANT: Hoskins, JoAnn
; APPLICANT: Wu, Chyun-Yeh Earnest
; TITLE OF INVENTION: Biosynthetic Gene Murd of Streptococcus
; TITLE OF INVENTION: pneumoniae
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Eli Lilly and Company
; STREET: Lilly Corporate Center
; CITY: Indianapolis
; STATE: Indiana
; COUNTRY: US
; ZIP: 46285
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/665,435A
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Webster, Thomas D
; REGISTRATION NUMBER: 39, 872
; REFERENCE/DOCKET NUMBER: X-9900
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 317-276-3334
; TELEFAX: 317-276-3861
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1350 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: mRNA
; HYPOTHETICAL: NO

```



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 14, 2003, 01:45:04 ; Search time 113 Seconds  
(without alignments)  
5663.766 Million cell updates/sec

Title: US-09-701-229-1  
Perfect score: 1450  
Sequence: 1 cgtgctgacgctcgcca.....tggtagcgcgacgagcgc 1450

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_NA: \*  
1: /cgn2.6/ptodata/1/ina/5A\_COMB.seq: \*  
2: /cgn2.6/ptodata/1/ina/5B\_COMB.seq: \*  
3: /cgn2.6/ptodata/1/ina/6A\_COMB.seq: \*  
4: /cgn2.6/ptodata/1/ina/6B\_COMB.seq: \*  
5: /cgn2.6/ptodata/1/ina/PCITUS\_COMB.seq: \*  
6: /cgn2.6/ptodata/1/ina/backfiles1.seq: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	1380.6	95.2	1401	4	US-09-252-991A-7861
C 2	1364.6	94.1	1371	4	US-09-252-991A-7702
C 3	567	39.1	567	4	US-09-252-991A-7787
C 4	438	30.2	564	4	US-09-252-991A-7928
C 5	353.6	24.4	360	4	US-09-252-991A-7552
C 6	258	17.8	1359	4	US-09-252-991A-7624
C 7	239	16.5	1416	4	US-09-328-353-1282
C 8	173.8	12.0	441529	3	US-09-103-840A-1
C 9	172.2	11.9	4403765	3	US-09-103-840A-2
C 10	149.8	10.3	1830121	4	US-09-557-884-1
C 11	149.8	10.3	1830121	4	US-09-643-990A-1
C 12	120	8.3	1170	4	US-09-252-991A-7929
C 13	76	5.2	9515	1	US-08-920-812-13
C 14	76	5.2	9515	1	US-08-920-827-13
C 15	76	5.2	9515	1	US-08-921-177-13
C 16	76	5.2	9515	1	US-08-362-577C-13
C 17	76	5.2	9515	2	US-08-920-828-13
C 18	73.6	5.1	44377	2	US-08-804-227C-7
C 19	73.6	5.1	44377	2	US-08-804-198-1
C 20	63.2	4.4	2325	4	US-09-252-991A-218
C 21	63.2	4.4	2784	4	US-09-252-991A-194
C 22	63.2	4.4	2946	4	US-09-252-991A-227
C 23	62.2	4.3	1335	4	US-09-252-991A-8206
C 24	62.2	4.3	1386	4	US-09-252-991A-8096
C 25	62	4.3	1287	4	US-09-252-991A-7860
C 26	62	4.3	5970	3	US-09-320-878-21
C 27	62	4.3	5970	4	US-09-141-908-11

C 28	62	4.3	5970	4	US-09-657-440-21	Sequence 21, Appl
C 29	61.4	4.2	1224	4	US-09-252-991A-11306	Sequence 11306, A
C 30	60.8	4.2	1248	3	US-09-105-537-7	Sequence 7, Appl
C 31	60.8	4.2	13613	3	US-09-105-537-3	Sequence 3, Appl
C 32	60.6	4.2	2721	6	5215881-2	Patent No. 5215881
C 33	60.6	4.2	8438	1	US-07-945-283-1	Sequence 1, Appl
C 34	60.2	4.2	1452	4	US-09-252-991A-3933	Sequence 3933, Ap
C 35	60.2	4.2	4284	4	US-09-252-991A-4006	Sequence 4006, Ap
C 36	59.6	4.1	15872	3	US-09-105-537-1	Sequence 1, Appl
C 37	59	4.1	801	2	US-08-934-481-3	Sequence 3, Appl
C 38	59	4.1	801	4	US-09-290-602-3	Sequence 3, Appl
C 39	59	4.1	2193	2	US-08-934-481-1	Sequence 1, Appl
C 40	59	4.1	2193	4	US-09-290-602-1	Sequence 1, Appl
C 41	58.8	4.1	1545	4	US-09-266-965-49	Sequence 49, Appl
C 42	58.8	4.1	53500	4	US-09-266-965-76	Sequence 76, Appl
C 43	58.6	4.0	1554	4	US-09-252-991A-14544	Sequence 14544, A
C 44	58.6	4.0	1761	4	US-09-252-991A-14820	Sequence 14820, A
C 45	58	4.0	1050	4	US-09-252-991A-13134	Sequence 13134, A

ALIGNMENTS

RESULT 1

US-09-252-991A-7861/c  
; Sequence 7861, Application US/09252991A  
; Patent No. 6551795

GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 7861  
; LENGTH: 1401  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-7861

Query Match	95.2%	Score	1380.6	DB	4	Length	1401
Best Local Similarity	99.7%	Pred. No.	9.4e-252				
Matches	1383	Conservative	0	Mismatches	4	Indels	0
Gaps	0						
QY	2	GTGCTGATCGGCTCGCCACCTTGAAGCTTGCCTTGAGGACGAGAGAGATGAGCTGAT	61				
Db	1387	GTGCTGATCGGCTCGCCACCTTGAAGCTTGCCTTGAGGACGAGAGAGATGAGCTGAT	1328				
QY	62	CGCCTCGACACTTCCGCATCGTTCGCTCGGCAAGAGCGGATTCCTCGTGGG	121				
Db	1327	CGCCTCGACACTTCCGCATCGTTCGCTCGGCAAGAGCGGATTCCTCGTGGG	1268				
QY	122	CTACCTGGCGCGCGGCTTGCCTTCGCTCGGCTGATACCGAGAACCCCGGA	181				
Db	1267	CTACCTGGCGCGCGGCTTGCCTTCGCTCGGCTGATACCGAGAACCCCGGA	1208				
QY	182	GCTGGCCACCTTGCCTGCGCAGTATCCGAGGTGGAAGTTCGCTGGCGGAATCGACGC	241				
Db	1207	GCTGGCCACCTTGCCTGCGCAGTATCCGAGGTGGAAGTTCGCTGGCGGAATCGACGC	1148				
QY	242	CGAGTTCCTCTCTCCGCCCGGAACTCTATCTAGCCCCCGCTTGTCTCGTGGCACCC	301				
Db	1147	CGAGTTCCTCTCTCTCCGCCCGGAACTCTATCTAGCCCCCGCTTGTCTCGTGGCACCC	1088				
QY	302	TCGGTGGTACAGGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG	361				
Db	1087	TCGGTGGTACAGGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG	1028				

```
QY 362 CCGGAGGCGAAGGCCCCGATCGTGGCCATCACCGGTTCCAAACGCGAAGACACCGTGAC 421
Db 1027 CCGGAGGCGAAGGCCCCGATCGTGGCCATCACCGGTTCCAAACGCGAAGACACCGTGAC 968
QY 422 CACCTGTGGGCGAATGGCGGTGGCGCGGACAAAGCGTGTCCGCGTGGCGGCAACCT 481
Db 967 CACCTGTGGGCGAATGGCGGTGGCGCGGACAAAGCGTGTCCGCGTGGCGGCAACCT 908
QY 482 CGGACCCCGCGCTCGACCTGCTGGCGGACGACATCGAGCTGTACGTTGGAGCTGTC 541
Db 907 CGGACCCCGCGCTCGACCTGCTGGCGGACGACATCGAGCTGTACGTTGGAGCTGTC 848
QY 542 GAGTTCCAGTGGAAACCTCGATCGCTCAACCGCGAGGTGGCGACCGTGTGTAACGT 601
Db 847 GAGTTCCAGTGGAAACCTCGATCGCTCAACCGCGAGGTGGCGACCGTGTGTAACGT 788
QY 602 CAGCGAAGACCATATGGATCGCTAGGAGCGCATGGCTGACTACCACTGGCCAAAGCCG 661
Db 787 CAGCGAAGACCATATGGATCGCTAGGAGCGCATGGCTGACTACCACTGGCCAAAGCCG 728
QY 662 GATCTTCCGCGTGGCCGACGAGTGGTGAATCGCGCCGATGCCCTGACCCGACCGCT 721
Db 727 GATCTTCCGCGTGGCCGACGAGTGGTGAATCGCGCCGATGCCCTGACCCGACCGCT 668
QY 722 GATCGCGATACCGTGGCGTCTGCTGGTTCGCGCTGAACAAGCCGACTTCAAGGCTTT 781
Db 667 GATCGCGATACCGTGGCGTCTGCTGGTTCGCGCTGAACAAGCCGACTTCAAGGCTTT 608
QY 782 CGGCTGATCCAGGAAGACGCGGAGTGGCTGCAACCTATTTCAACCGCTGCGCGCGTGGC 841
Db 607 CGGCTGATCCAGGAAGACGCGGAGTGGCTGCAACCTATTTCAACCGCTGCGCGCGTGGC 548
QY 842 GGTGGGCACTGAAGTCCGTGGCGCCCAACTATTTCAACCGCTGCGCGCGTGGC 901
Db 547 GGTGGGCACTGAAGTCCGTGGCGCCCAACTATTTCAACCGCTGCGCGCGTGGC 488
QY 902 GCTGGGCACTGGTGGCTGCGCTGCGTTCGAGCGCATGCTGGCGCGTGAAGGCGTTTC 961
Db 487 GCTGGGCACTGGTGGCTGCGCTGCGTTCGAGCGCATGCTGGCGCGTGAAGGCGTTTC 428
QY 962 CGGCTGGCTCATCGCTGCAAGTGGTACGCGAGCGCGGCGTGAAGTACGACG 1021
Db 427 CGGCTGGCTCATCGCTGCAAGTGGTACGCGAGCGCGGCGTGAAGTACGACG 368
QY 1022 TTCAAAGGCCAACACGTCGCGCGCCCTGGCGCGATCGAGGGGCTGGGTGCGGACAT 1081
Db 367 TTCAAAGGCCAACACGTCGCGCGCCCTGGCGCGATCGAGGGGCTGGGTGCGGACAT 308
QY 1082 CGAGCGCAAGCTGGTGGCTGCTCGCGCGGCGGAGAGCGGCAAGGGCGCCGATTCATGACCT 1141
Db 307 CGAGCGCAAGCTGGTGGCTGCTCGCGCGGCGGAGAGCGGCAAGGGCGCCGATTCATGACCT 248
QY 1142 CGCGAGCGGCTCGCGGCTTCTCGCGGCGGTTGTTACTGCTGGCGGCTGACGCGCGGCT 1201
Db 247 CGCGAGCGGCTCGCGGCTTCTCGCGGCGGTTGTTACTGCTGGCGGCTGACGCGCGGCT 188
QY 1202 GATTGCCAGGCACTGGGCAACCGGTTACCGCTGGTGGCGGCTCGCAACGCTGGACGAAGC 1261
Db 187 GATTGCCAGGCACTGGGCAACCGGTTACCGCTGGTGGCGGCTCGCAACGCTGGACGAAGC 128
QY 1262 AGTCGCGAGCGCGCGGCTGGCGCGGAGGCGATGCGGTGCTGTTGCGCGGCGCTG 1321
Db 127 AGTCGCGAGCGCGCGGCTGGCGCGGAGGCGATGCGGTGCTGTTGCGCGGCGCTG 68
QY 1322 CGCGAGGCTGGACATGTTCAAGAACTTCGAAGAACGCGGAGCGGCTGTTGCCCAAAGCCGT 1381
Db 67 CGCGAGGCTGGACATGTTCAAGAACTTCGAAGAACGCGGAGCGGCTGTTGCCCAAAGCCGT 8
QY 1382 AGAGGAG 1388
Db 7 AGAGGAG 1
```

```
RESULT 2
US-09-252-991A-7702
; Sequence 7702, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 7702
; LENGTH: 1371
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-7702

Query Match 94.18; Score 1364.6; DB 4; Length 1371;
Best Local Similarity 99.7%; Pred. No. 9.6e-249;
Matches 1367; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 27 AGCTGCGCTTGAGGACGAGAGATGAGCTGATCGCTCCGACCACTTCCGCGATGTT 86
Db 1 AGCTGCGCTTGAGGACGAGAGATGAGCTGATCGCTCCGACCACTTCCGCGATGTT 60
QY 87 GTCGGCTCGGCAAGAGCGCATGTCCCTGGTGGCTTACCTGGCGCGCCGCGGCTTCCCT 146
Db 61 GTCGGCTCGGCAAGAGCGCATGTCCCTGGTGGCTTACCTGGCGCGCCGCGGCTTCCCT 120
QY 147 TTGCGCTGTCGATACCCGAGAGAACCCCGCGGAGTGGCCACCTTGGTGCCAGTAT 206
Db 121 TTGCGCTGTCGATACCCGAGAGAACCCCGCGGAGTGGCCACCTTGGTGCCAGTAT 180
QY 207 CCGCAGGTGGAAGTGGCTTGGCGCAACTCGAGCGGAGTTCCTCTCTCGCCCGCGAA 266
Db 181 CCGCAGGTGGAAGTGGCTTGGCGCAACTCGAGCGGAGTTCCTCTCTCGCCCGCGAA 240
QY 267 CTCTATGTACGCCCGGCTTGTGCTGCGCAACCCCTGCGTGTGTACAGGCGCGCGGAAA 326
Db 241 CTCTATGTACGCCCGGCTTGTGCTGCGCAACCCCTGCGTGTGTACAGGCGCGCGGAAA 300
QY 327 GCGTGGCATCTCCGCTGACATCTCTTGGCGCGAGGCGGAGGCGCGATGCTC 386
Db 301 GGTGTGGCATCTCCGCTGACATCTCTTGGCGCGAGGCGGAGGCGCGATGCTC 360
QY 387 GCCATCACCGGTTCCAAACGCAAGAGACCGCTGACCACTTGGTGGCGGAAATGGCGGTG 446
Db 361 GCCATCACCGGTTCCAAACGCAAGAGACCGCTGACCACTTGGTGGCGGAAATGGCGGTG 420
QY 447 GCGCGGACAAAGCGTGTGCGCGTGGCGGCAACCTTGGGACCCCGCGGCTGACCTGCTG 506
Db 421 GCGCGGACAAAGCGTGTGCGCGTGGCGGCAACCTTGGGACCCCGCGGCTGACCTGCTG 480
QY 507 GCGGACGACATCGAGCTGTACGTTGGAGCTGTTCAGCTTCCAGCTTGGAAACCTTGGAT 566
Db 481 GCGGACGACATCGAGCTGTACGTTGGAGCTGTTCAGCTTCCAGCTTGGAAACCTTGGAT 540
QY 567 CGCCTCAACCGCGAGGTGGCGACCGTGTGTAAGCTCAGCGAAGACCATATGATGATGCTAC 626
Db 541 CGCCTCAACCGCGAGGTGGCGACCGTGTGTAAGCTCAGCGAAGACCATATGATGATGCTAC 600
QY 627 GAGGGATGGCTGACTACCACTTGGCGAAGCGGATCTTCCGCGGTGCCCGCAGGTC 686
Db 601 GAGGGATGGCTGACTACCACTTGGCGAAGCGGATCTTCCGCGGTGCCCGCAGGTC 660
QY 687 GTGTTGAATCGCGCGGATGCGCTGACCGACCGCTGATCGCGGATACCGTGGCGTGTG 746
Db 661 GTGTTGAATCGCGCGGATGCGCTGACCGACCGCTGATCGCGGATACCGTGGCGTGTG 720
```



QY 747 TCGTTCGGCCTGAACAGCGCGGACTTCAAGGCTTTCAGGCTTTCGGCCTGATCGAGGAAGACGGCCAG 806  
Db 721 TCGTTCGGCCTGAACAGCGCGGACTTCAAGGCTTTCGGCCTGATCGAGGAAGACGGCCAG 780  
QY 807 AAGTGCCTGGCCTCCAGTTCACAGCTGCTGCGGCTTGGGGAAGTCCGCTGCG 866  
Db 781 AAGTGCCTGGCCTCCAGTTCACAGCTGCTGCGGCTTGGGGAAGTCCGCTGCG 840  
QY 867 GCCACAACTATTCAACAGCGCTCGCGCGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGG 926  
Db 841 GCCACAACTATTCAACAGCGCTCGCGCGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGG 900  
QY 927 TTCGAGCCCATGCTGCGGCGCTGAAGGCGTTTTCGGGCTTTCGGGCTTTCGGGCTTTCGGG 986  
Db 901 TTCGAGCCCATGCTGCGGCGCTGAAGGCGTTTTCGGGCTTTCGGGCTTTCGGGCTTTCGGG 960  
QY 987 GTACGGAGCGCGAGCGCTGAGCTACTACGAGCTTCAAGGCTTTCAGGCTTTCAGGCTTTCAGG 1046  
Db 961 GTACGGAGCGCGAGCGCTGAGCTACTACGAGCTTTCAGGCTTTCAGGCTTTCAGGCTTTCAGG 1020  
QY 1047 GCGCTGGCGGCTGAGGCGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGG 1106  
Db 1021 GCGCTGGCGGCTGAGGCGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGG 1080  
QY 1107 GCGGAGAGCGCAAGGCGCGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGG 1166  
Db 1081 GCGGAGAGCGCAAGGCGCGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGG 1140  
QY 1167 GCGGCGCTGCTGCTGCGGCTGACGCGCGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCT 1226  
Db 1141 GCGGCGCTGCTGCTGCGGCTGACGCGCGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCT 1200  
QY 1227 GTACCGCTGCTGCGGCTGCAACGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCT 1286  
Db 1201 GTACCGCTGCTGCGGCTGCAACGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCT 1260  
QY 1287 CGGGAAGCGATGCGGCTGCTGCTGCGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGG 1346  
Db 1261 CGGGAAGCGATGCGGCTGCTGCTGCGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGG 1320  
QY 1347 TTCGAAGACGCGGAGCGCTGCTGCGCAAGCGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGG 1397  
Db 1321 TTCGAAGACGCGGAGCGCTGCTGCGCAAGCGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGG 1371

RESULT 3  
US-09-252-991A-7787/c  
; Sequence 7787, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; PRIOR FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 7787  
; LENGTH: 567  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-7787

Query Match 39.1%; Score 567; DB 4; Length 567;  
Best Local Similarity 100.0%; Pred. No. 1,3e-98;  
Matches 567; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 772 TCAAGGCTTTCGGCCTGATCGAGGAAGACGGCCAGAGTGGCTGGCTTCCAGTTCGACA 831

Db 567 TCAAGGCTTTCGGCCTGATCGAGGAAGACGGCCAGAGTGGCTGGCTTCCAGTTCGACA 508  
QY 832 AGCTGCTGCGGTTGGCGAACTCAAGATCCGTGGCGCCACAACTATTCCAAAGCGCGCTCG 891  
Db 507 AGCTGCTGCGGTTGGCGAACTCAAGATCCGTGGCGCCACAACTATTCCAAAGCGCGCTCG 448  
QY 892 CCGCGCTGCGGTTGGCGCCATGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGG 951  
Db 447 CCGCGCTGCGGTTGGCGCCATGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGG 388  
QY 952 AGCGCTTTCGGCGCTTGGCTCATCGCTGCCAGTGGGTACGCGAGCGCGAGGGGCTGAGCT 1011  
Db 387 AGCGCTTTCGGCGCTTGGCTCATCGCTGCCAGTGGGTACGCGAGCGCGAGGGGCTGAGCT 328  
QY 1012 ACTACGACGATTCCAAAGCGCACCAACGCTGCGGCGCCCTGCGGCGGATCGAGGGGCTGG 1071  
Db 327 ACTACGACGATTCCAAAGCGCACCAACGCTGCGGCGCCCTGCGGCGGATCGAGGGGCTGG 268  
QY 1072 GTGCCGACATCGACGCAAGCTGCTGCTGCTGCGGCGGAGAGCGAGGCGCGGCTGTT 1131  
Db 267 GTGCCGACATCGACGCAAGCTGCTGCTGCTGCGGCGGAGAGCGAGGCGCGGCTGTT 208  
QY 1132 TCCATGACCTGCGGAGCGGCTGCGGCGCTTCTGCGGCGGCTGCTGCTGCTGCTGCTGCTG 1191  
Db 207 TCCATGACCTGCGGAGCGGCTGCGGCGCTTCTGCGGCGGCTGCTGCTGCTGCTGCTGCTG 148  
QY 1192 AGCGCGGCTGATTGCCAGGCACTGGGCAACGCGTACCGCTGCTGCGGCTGCGCAACGC 1251  
Db 147 AGCGCGGCTGATTGCCAGGCACTGGGCAACGCGTACCGCTGCTGCGGCTGCGCAACGC 88  
QY 1252 TGGACGAACAGTCCGCGAGCGCGGCTGCGGCTGCGGCGGATGCGGCTGCTGCTGCTG 1311  
Db 87 TGGACGAACAGTCCGCGAGCGCGGCTGCGGCTGCGGCGGATGCGGCTGCTGCTGCTGCTG 28  
QY 1312 CGCGCGCTTGGCGGAGCCTTGGACATGT 1338  
Db 27 CGCGCGCTTGGCGGAGCCTTGGACATGT 1

RESULT 4  
US-09-252-991A-7928/c  
; Sequence 7928, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; PRIOR FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 7928  
; LENGTH: 564  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-7928

Query Match 30.2%; Score 438; DB 4; Length 564;  
Best Local Similarity 100.0%; Pred. No. 2.4e-74;  
Matches 438; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1013 CTACGACGATTCCAAAGCGCACCAACGCTGCGGCGCCCTGCGGCGGATCGAGGGGCTGG 1072  
Db 564 CTACGACGATTCCAAAGCGCACCAACGCTGCGGCGCCCTGCGGCGGATCGAGGGGCTGG 505  
QY 1073 TCCCGACATCGAGCGGAAGCTGCTGCTGCTGCGGCGGAGAGCGGCAAGGGCGCCGCTTT 1132  
Db 504 TCCCGACATCGAGCGGAAGCTGCTGCTGCTGCGGCGGAGAGCGGCAAGGGCGCCGCTTT 445

Query Match	24.4%	Score 353.6;	DB 4;	Length 360;
Best Local Similarity	98.9%;	Pred. No. 1.8e-58;		
Matches 356;	Conservative 0;	Mismatches 4;	Indels 0;	Gaps 0;
QY	61	TCGCCCTCCGACCACTTCCGCATCGTTGTGCGGCCTCGGCAAGACGGGATGTCCTGGTGC	120	
Db	1	TCGCCCTCCGACCACTTCCGCATCGTTGTGCGGCCTCGGCAAGACGGGATGTCCTGGTGC	60	
QY	121	GCTACTTGGCGCGCGGGCTTGCCCTTTCGCCGTGTGCGATACCCGAGAGAAACCCGCCGG	180	
Db	61	GCTACTTGGCGCGCGGGCTTGCCCTTTCGCCGTGTGCGATACCCGAGAGAAACCCGCCGG	120	
QY	181	AGTGGCCACCCCTGCGTGCCAGTATCCGAGGTGGAAGTGCCTTGGCGGGAACCTCGACG	240	
Db	121	AGTGGCCACCCCTGCGTGCCAGTATCCGAGGTGGAAGTGCCTTGGCGGGAACCTCGACG	180	
QY	241	CCGAGTTCCTCTGCTCCGCCCGGGAACCTATATGTACGCCCCGGCTGTGCGTCGCGACCC	300	
Db	181	CCGAGTTCCTCTGCTCCGCCCGGGAACCTATCTACGCCCCGGCTGTGCGTCGCGACCC	240	
QY	301	CTGCGCTGTGTACAGGCGCGCGGAAGGGGTGCGCATCTCCGGTGACATCGATCTCTTCG	360	
Db	241	CCGCGCTGTGTACAGGCGCGCGGAAGGGTGTGCGCATCTCCGGTGACATCGATCTCTTCG	300	
QY	361	CCGCGAGGCGAAGGCCCGGATGTCGCCATCACCGGTTCCCAACGCGGAAGACCCGTTGA	420	

	Query Match	17.8%	Score 258;	DB 4;	Length 1359;
	Best Local Similarity	100.0%;	Prod. No. 1.9e-40;		
	Matches 258;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1193	CGCGGGCTGATTGCCAGGCACTGGGCAACGGGTACCGCTGTCGGCGCTGCACAGGCT	1252		
DB	1	CGCGGGCTGATTGCCAGGCACTGGGCAACGGGTACCGCTGTCGGCGCTGCACAGGCT	60		
QY	1253	GGACGAAGCAAGTCCGGGAGCGCCGCGAGCTGGCCCCGGGAAAGGCGATCGGTGTGTGTC	1312		
DB	61	GGACGAAGCAAGTCCGGGAGCGCCGCGAGCTGGCCCCGGGAAAGGCGATCGGTGTGTGTC	120		
QY	1313	CGCGGGCTGGCGGAGCCTGGACACTGTTCAAGAAGTTCGAAGAAGCGGAGCGCCTGTCGC	1372		
DB	121	CGCGGGCTGGCGGAGCCTGGACACTGTTCAAGAAGTTCGAAGAAGCGGAGCGCCTGTCGC	180		
QY	1373	CAAAAGCGGTAGAGGAGCTACCGTGTATGCTGTCGGTGTTCGGCCCCCTCCCGTCCGCGCTG	1432		
DB	181	CAAAAGCGGTAGAGGAGCTACCGTGTATGCTGTCGGTGTTCGGCCCCCTCCCGTCCGCGCTG	240		
QY	1433	TTGAGCGGCGACGGGCATC	1450		
DB	241	TTGAGCGGCGACGGGCATC	258		

Query Match	16.5%	Score 239;	DB 4;	Length 1416;
Best Local Similarity	51.6%;			
Pred. No. 7.3e-37;				
Matches 604; Conservative	0;	Mismatches 555;	Indels 12;	Gaps 2;

QY 213 GTGGAAGTGGTTGGCGGCACTGAGCGCGAGTTCCTCTGCTCCCGCCGCGAAGTCTAT 272  
Db 220 GTTAAACCAGTTTGTGCTAGCTGTATCAAGAAATATTATTAAAGCAAGAAAGATATT 279  
QY 273 GTGAGCGCGGCTGTGCTGCGACCCCTGCTGCGCTGATACAGCCCGCGGAAAGCGTG 332  
Db 280 TTAAGCCAGGCTTGCACCAATATACCGAAATTCAGGACGCTATTGCTAAAGTATT 339  
QY 333 CGCATCTCGGTGACATCGATCTCTTCGCCGCGAGGCGAAGCCCGCATCGTCCGCATC 392  
Db 340 TCTGTGTGGCGGATATCCAGTTATTGCGCGTGCCTGATGTCGGGATTTGGCGATT 399  
QY 393 ACCGGTTCCAAACGGAAGACACCGCTGACACCCCTGGTGGGAAATGGCGGTGGCGG 452  
Db 400 ACAGGTTCCAAATCAAAAAGTACTGTAACACCTTTAATGGCTTAATGGCTAAGATGA 459  
QY 453 GACAAAGTGTCTGCGCTGCGGCGCAACCTCGGCACCCCGCGCTGACCTGTGGCGGAC 512  
Db 460 GGTAAGAAATGTCAGTAGCGGCAACCTTGTGCGACGAGTGTAGTTACTTTAAAGAT 519  
QY 513 GACATCGAGTGTACGTGTGGAGCTGTGAGCTTCCAGCTGGAACCTGCGATCGCCTC 572  
Db 520 CAACCAAGAGTACTGTCTGAGTTATCAAGCTTTCAGTTAGAACCAACCTCTCATT 579  
QY 573 AAGCGGAGTGGCGACCGTGTGAAGCTGACGGAAGACCATATGATCGCTACGACGGC 632  
Db 580 AAGCTGAGGTAGCAGTGTCTTCAATATGAGTGAAGACATTTAGACGCTCATGAAAT 639  
QY 633 ATGGCTGACTTACCACCTGCGCAAGCACCGATCTCCGCGGTGCGCGCAGCTCGGTG 692  
Db 640 ATGCTGGGTATCATCAACAAACATCATGTTTCCAAAGGCGCTAAAAGATGTGATT 699  
QY 693 AATCGCGCGATGCCGTGACCGGCTGATCGCGATACCGTGCCTGCTGCTGCTTC 752  
Db 700 AACCGAGATGATGCGTTAAGCGCTCCACTTGTTCAGATACGACACCATGCAAGCTT 759  
QY 753 GGCCTCAACAGCGGACTTCAAGGCTTTCGCG ---CTGATCGAGGAAGCGGCGAAG 809  
Db 760 GGTTTAAGCGCACCGGATTTAATCAATATGCGTTTAAAGAGTGCACCGGTACGCT 819  
QY 810 TGCTGGCTTCCAGTTCGACAGCTGCTGCGGTGGGAACTGAAGATCGTGGCGGC 869  
Db 820 TGCTTGTCTGTTTACAGCGGTTAATAAAGCTCAGATTTATATTCAAGTATG 879  
QY 870 CACAATATTCAACGCGCTCGCGGCTGCGGCTGGCCATGCGGTGCGGCTGCGGCTTC 929  
Db 880 CACAATGTAGCTAATGCTTTAGCTTTAGCATTTGGGGGAAGCAATTTGGTTTACCTATG 939  
QY 930 GAGCGCATCTCGCGCGCTGAGGCGTTTTCGCGCTGGCTTCATCGCTGCGAGTGGTA 989  
Db 940 GAGTCAATGCTTGAACCTTTAAAAAATTTAAAGGACTAGACCGCTGTGAGTACGTT 999  
QY 990 CGCGAGCGGAGGGGTGAGCTACTACGACGATTCGAAGCCCAACCACTGCGGCGCGC 1049  
Db 1000 AAAACGTCATGATGCTGTTTATATATGACTCTAAGAGTACTAATGTTGCTGTACA 1059  
QY 1050 CTGGCGGCGATCGAGGCGCTGGTCCGACATCG -----ACGGCAAGTGGTGTG 1100  
Db 1060 CTTGACGAATTTGATGGCTTAGGTGCTGCATTGAAGTGAAGTGAAGTGAAGTTCACCTT 1119  
QY 1101 CTCGCGCGGAGAGCGCAAGGCGCGGATTTCCATGACCTGCGGAGCGGCTGCGGCG 1160  
Db 1120 ATTTTAGTGGACAGGGAAGGCAAGATTTTTCCTTTTACGTTCTTCTATCGAAAAA 1179  
QY 1161 TTCTGCGGCGGTGCTACTGCTGTGCGGTGACGCGGGCTGATTTGCCAGGCACTGGGC 1220  
Db 1180 TATGCCAAGTTGTGATTGATTGTTGAGATGCGCTGTCTGATCGAACAGCAATTCAT 1239  
QY 1221 AACGGGTACCGCTGGTGGCTGCGCAAGCTGGAAGCAAGCTCGGCGAGCGCGCGAG 1280  
Db 1240 GCGCAACATAAATTTTACATCAGCAACGCTTAAAGAGCTGTGCGAGCTGTGCAAGT 1299  
QY 1281 CTGGCGCGGAGGCGATGCGGTGCTGTGTCGCGCGGCTGCGCGAGCGCTGGACATGTC 1340

Db 1300 GAAACAAGCTGAAGATGGTATTGCTATCACCAGCATGTCAAGTTTGTATATGTTT 1359  
QY 1341 AGAAGCTTCGAGAAGCGGACCCCTGTTG 1371  
Db 1360 AAAAGTTAATGACCGTGTGTCAGCAGTTTG 1390

RESULT 8  
US-09-103-840A-1/C  
; Sequence 1, Application US/09103840A  
; Patent No. 6294328  
; GENERAL INFORMATION:  
; APPLICANT: FLEISCHMAN, Robert D.  
; APPLICANT: WHITE, Owen R.  
; APPLICANT: PRASER, Claire M.  
; APPLICANT: VENTER, John C.  
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM  
; TITLE OF INVENTION: TUBERCULOSIS  
; FILE REFERENCE: 24366-20007.00  
; CURRENT APPLICATION NUMBER: US/09/103,840A  
; CURRENT FILING DATE: 1998-06-24  
; NUMBER OF SEQ ID NOS: 2  
; SOFTWARE: Patentin Ver. 2.1  
; SEQ ID NO 1  
; LENGTH: 4411529  
; TYPE: DNA  
; ORGANISM: Mycobacterium tuberculosis  
; OTHER INFORMATION: H37Rv  
US-09-103-840A-1

Query Match 12.0%; Score 173.8; DB 3; Length 4411529;  
Best Local Similarity 51.7%; Pred. No. 2.3e-24;  
Matches 499; Conservative 0; Mismatches 427; Indels 39; Gaps 3;

QY 273 GTGAGCGCGGTGTGCTGCGCACCCCTGCTGAGCGGCTTACAGCCCGCCGCAAGCGGTG 332  
Db 2416179 GCGAGTCCCGGCTTCTCGCCGCAACCCGCTACTGGCGCGCGCGCGCGGGGTG 2416120  
QY 333 CGCATCTCGGTGACATCGATCTCTTCGCGCG -----CGAGGCG 371  
Db 2416119 CCGATCTGGGTGAGCTGGAGTTAGCTGGCGGTAGACGACGCGGCTGTACGAGCG 2416060  
QY 372 AAGGCGCGCATGCTGCGCCATCACCGTTCACGGAAGACGACCGTACACCGCTGCTGGTG 431  
Db 2416059 CCGCGCAGCTGCTGCTGCTGAGCGGACCAAGCGAAGACCAACGACGCTGATGCTG 2416000  
QY 432 GCGAAATGGCGGTGGCGCGGACAAAGCTGTGCGCGTGGCGGCAACCTCGCACCGCG 491  
Db 2415999 CACGCCATGCTGATCGCGGTGGCGCGCGCTGCTGTGGGCAATATCGSCAGTGG 2415940  
QY 492 GCGCTCGACTGCTGGCGGACGACATCGAGCTGTAGTGTGAGCTGTGAGCTTCAGT 551  
Db 2415939 GTGCTGGATGCTGAGACGAGCGCGGCGAGCTGTGCGCGTGGAGTTGTCCAGTTCCAG 2415880  
QY 552 CTGGAACCTGCGATCGCTCAACCGCGAGGTGGCGACCGTCTGAACGTCAGCGAAGAC 611  
Db 2415879 CTGCACTGGCGCGCTGCTGCTGCGCGGAGCGCGGCTGCTCAACATTGCCAAGAC 2415820  
QY 612 CATATGATGCTGACGAGCGATGCTGACTACACCTGCGGCGCAAGCACCGGATCTTCGCG 671  
Db 2415819 CACCTGGACTGGCATGCCAGATGGCGGAATACCGGCGCAAGCGCGGTGCTGACC 2415760  
QY 672 GGTGCGCGCGCTGCTGTGTAATCGCGCGATGCCCTGACCCGCGCTGATCCCGAT 731  
Db 2415759 GCGGGGTGAGCGTGGCGGGCTGGATGACCGGAGCGCGCGACTGCTGACGCGCTCA 2415700  
QY 732 ACCGTGCGGTGCTGCTGCG ---TTCGGCTGAAACAGCGGACTTCAAGGCTTTCGCGCTG 788  
Db 2415699 CCGGCGAGGTGCGGCTGCGCTTCCGGCTGCGGAGCGCGCGCGGAACTGGCGGTG 2415640  
QY 789 ATCGAGGAGAGCGGCGAGAGTGGCTGCTGCTTCCAGTTCCAGCAAGCTGCTGCGGTGCG 848



Query Match	Score	DB 4;	Length
10.38;	Score 149.8;	DB 4;	Length 1830121;

```

Best Local Similarity 51.0%; Pred. No. 7.2e-20;
Matches 440; Conservative 0; Mismatches 407; Indels 16; Gaps 3;

QY 272 TGTACGCCCCGGTGTGCTGCGCAGCCCTGCGCTGTGACAGCCCGCGGAAAGCGGT 331
DB 1203341 TATTAGCCCGGGTGGGTAAACACACAGAAATTAACACCGCACTTAAGCGGGAGT 1203400
QY 332 CGCATCTCCGGTGACATCGATCTCTTCGCCCGGAGGCGGAGCCCGCATCGTCGCCAT 391
DB 1203401 GGAAGTAATCGCGATATTGAATTATCTGCGCGCAGCAGCAAGCAATTTGGGGAT 1203460
QY 392 CACCGGTTCCAAAGGAGACACCGGTGACACCCCTGTGTGGGCGAAATGCGGTGGCGGC 451
DB 1203461 TACAGGTTCAATGGTAAAGTACCGTAACCTACTTATAGTTATGAAATGGCGAAAGTGC 1203520
QY 452 GGACAAGCGTGTGCGCGTGGCGGCGCAACCTCGGCAACCCCGCTGACACTGCTGGCGGA 511
DB 1203521 TGGTGTGAAAGTTGGTATGGCGGGAATATTGGGATTCGCCCTTTGTCATTTGAATGA 1203580
QY 512 CGACATCGAGCTGTACGTTGGAGCTGTGAGCTTCCAGCTTCCAGTGGAAACCTCGATCGCCT 571
DB 1203581 AGATTGTGAACCTTTATGTACTAGAGCTTTCTAGTTTTCAGCTTCGAGACAACTTATAGCTT 1203640
QY 572 CAACCCGAGTGGCGGCGCTGTGAGCTGAGCGAAGACCAATATGATGCTCTAGCAGCG 631
DB 1203641 AAAAGCTCGCGCAGCGACTGCTTTGAACGTGACTGAAGATCATATGATGCTATATGGA 1203700
QY 632 CATGGCTGACTTACCACTGGCGCAAGCAGCCGATCTTCGGCGTCCCGCCAGGTCGTGGT 691
DB 1203701 TTTAGAGATTTATGCCAAGCAAAATTTACGATTTTATCATATGCTAAAGTAGTGTGTT 1203760
QY 692 GAATCGGCGGATCCCTGTGACCCGAGCGCTGATCGCGAT--ACCGTCCGCTGTGCTGTCG 749
DB 1203761 GAACAATGAAGATAGGCTGACTTTTGGGAAACGAAATCAAGCAAAATACCGTTTC 1203820
QY 750 TTCGCGCTGACAGCGCGACTTCAAGCTTTCGGCTGTGTCGAGGAGACGCGGCGGCAAG 809
DB 1203821 TTTTGGGAAATAGTCGGGATTTATGGCTAAAAACT-----GAAATGGCAAGCAA 1203872
QY 810 TGGCTGGCGTCCAGTGTGCAAGCTGCTGCGGTTGCGCAACTGAAGATCCGTCGCGCC 869
DB 1203873 TATTATGTAAGATGAAGTATTTTACCTTGTGAAGAGCTACATGTTGGTGGCG 1203932
QY 870 CACAATATTCCAAAGCGCTGCGCGCTGCGCTGCGGATGCGGCTGCGGCTGCGGCTTC 929
DB 1203933 CATAATTATATGACATTTTGGCAGCAACAGCATTTGCAAGATAGTATTAATTA 1203992
QY 930 GAGCCATGCTCGCGCGCTGAAGCGCTTTTCGGCTGCTCATCCTGCCAGTGGTA 989
DB 1203993 GATTCAATTCGTACCGCACCTGCTCATTTCAAAGGGTTAGATCATCCTTTTCAATTAGT 1204052
QY 990 CGGAGCGGCGGCGGTGAGCTACTAGCAGGATTCACAGGCAACCAAGTCCGCGCGCC 1049
DB 1204053 CATCAAGCTAATGCACTTCGTTGATTAATGACTTAAGCAACAATGTGGGAGTACA 1204112
QY 1050 CTGCGCGGATCAGAGGCGGTGGTGGCGCATCGACGCAAGTGTGCTGTCGCGCGC 1109
DB 1204113 GTTCTGCTATTTGCTGGGCT-----TTATATTAGGGTAAATTTGCTATTGCTAGGC 1204166
QY 1110 GGAGACGGAAGGCGCGGATTT 1132
DB 1204167 GGAGACGGAAGGCGCTGATTT 1204189

```

```

RESULT 12
US-09-252-991A-7929/c
; Sequence 7929, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136

```

```

; CURRENT APPLICATION NUMBER: US/09/252.991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; LENGTH: 1170
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-7929

Query Match      8.3%; Score 120; DB 4; Length 1170;
Best Local Similarity 100.0%; Pred. No. 1.8e-14;
Matches 120; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GTGCTGATCGCGCTCGCCACCTTGAAGCTGCGTTGAGGACGAAGAGCATGAGCCTGAT 61
DB 120 GTGCTGATCGCGCTCGCCACCTTGAAGCTGCGTTGAGGACGAAGAGCATGAGCCTGAT 61
QY 62 CGCCTCCGACCACTTCCGCATCGTTGCGGCTCGGCAAGAGCGGCATGTCCTGGTGGC 121
DB 60 CGCCTCCGACCACTTCCGCATCGTTGCGGCTCGGCAAGAGCGGCATGTCCTGGTGGC 1

RESULT 13
US-08-920-812-13
; Sequence 13, Application US/08920812
; Patent No. 5763188
; GENERAL INFORMATION:
; APPLICANT: Ohno, Tsuneo
; APPLICANT: Matsuhisa, Akio
; APPLICANT: Uehara, Hirotsugu
; APPLICANT: Eda, Soji
; TITLE OF INVENTION: Probe for Diagnosing Infectious Disease
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 6300 Sears Tower, 233 South Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States of America
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/920,812
; FILING DATE: 29-AUG-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/362,577
; FILING DATE: 27-MAR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Rin-Laures, Li-Hsien
; REGISTRATION NUMBER: 33,547
; REFERENCE/DOCKET NUMBER: 19036/32420
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312/474-6300
; TELEFAX: 312/474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 9515 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
; ORIGINAL SOURCE:
; ORGANISM: Pseudomonas aeruginosa

```

STRAIN: Clinical Isolate P2-2  
US-08-920-812-13

Query Match 5.2%; Score 76; DB 1; Length 9515;  
Best Local Similarity 44.3%; Pred. No. 4e-06;  
Matches 455; Conservative 0; Mismatches 560; Indels 13; Gaps 3;

QY 283 GCTTGTGCTGCGCACCCTGCTGTATACAGGCCCGCGAAGGCGTGCATCCG 342  
DB 4334 GCTTCCAGCGGACACCTTGTCCATCGCAAGGCGCTGAGCTCGGCTACATCCCATG 4393  
QY 343 GTGACATCATCTCTTCGCGCGGAGGGAAGGCCCGGATCGTCGCCATCACCGGTTCCA 402  
DB 4394 GCGGCTGTACTTCGCAAGCGCATCGCGAGGTGTGTGGAGAGGCGGGGTTCG 4453  
QY 403 ACCGGAAGAGCAGCACCTGTGTGGCGAAATGCGGTTGGCGGACAAACGCTG 462  
DB 4454 CCACGGCTGACCTATTCCGCGCACCCGCGGCGGCTGGCCATCGCCAACTCA 4513  
QY 463 TCGCGG-TCGGGGCAACCTCGGCACCCCGGCGCTCGACCTGTCGCCGACGACATCGAG 521  
DB 4514 AGGCTCGGAGCAGGCGGTGTACGCGGGTTCAGGAGAGACCGGCCCTACCTGCAA 4573  
QY 522 CTGTAGCTGTGAGCTGTGAGCTTCCAGCTGGAACCTGCGATCGCTCAACGCGAG 581  
DB 4574 CGTGTCTGCGGAGGTCTTCGGCGACCATCGCTGTGCGGAGGTCCAGGCGCGCGC 4633  
QY 582 GTGGGACCTGTGTGAGCTGACGGAAGACCATATGATGCTGCTACGACGCGATGGCTGAC 641  
DB 4634 TTGCTGCGCGCTGAGTTCGCGGAGGACAGGTACCGCGAAGCGCTTCGCCAACGAG 4693  
QY 642 TACCACCTGCCAAGCAGCGGATCTTCGCGGTGCCCGCAGGTGCTGTGAATCGCGC 701  
DB 4694 AAGCATCTGCGCTGCGCTGCCACCATCGCGGGTTCGAGAGGCGGTGATCATCGC 4753  
QY 702 GATGCCCTGACCGACCGCTGATCGCGGATACCGTCCGCTGCTGTGCTTCGCGCTGAC 761  
DB 4754 TCACCCCTGCGCGCATGATCATGCGCGCGCGCTGTGCGCGGCGGTGCCGAGATCGAC 4813  
QY 762 AAGCCGACTTCAAGCGCTTTCGCGCTGATCGAGGAAGCGGCGAGGCTGCGGCTTC 821  
DB 4814 GAATGATCGACAAGAC-----CCGTATCGCGGTGATCGCACCGCGCGAGATCGGC 4867  
QY 822 CAGTTGCAAGAGTGTGCGGTTGGCGAACTGAAGATCGGTGGCGCGCCACAACTATTC 881  
DB 4868 GTGCTGTAGCGCGCGCGCGCGCGCTCGCGCGGTGCGCTGCGACACGAGGCTCC 4927  
QY 882 AAGCGCTCGCGCGCTGGCGTGGCCATGCGGTGCGGCTGCGCTCGACGCCATGCTC 941  
DB 4928 CCCATACGACGATGCGCGCGCTGGCGACCGCGCGGAACGTTTCGCGCTCTGGCGG 4987  
QY 942 GCGCGCTGAAGCGCTTTCGCGCTGCTCATCGCTGCGGTGCGGTACGCGAGCGGAC 1001  
DB 4988 CAATGCTAAGCAACATCACACATGCCATCGCTGTGGAGGTGTTCCATGTTCAAG 5047  
QY 1002 GCGGTGAGTCTACGACGATTCGAAGGCGCAACAGTGTGCGCGCGCGCTGCGCGGATC 1061  
DB 5048 TCCTTGACCAAGTACGACACGCTGTTTCCCGGTGCTGCTGCTGCTGCGGTTTCGCC 5107  
QY 1062 GAGGGCTGGTCCGACATCGACGCAAGCTGTGCTGCTCCCGCGGAGAGCGGCAAG 1121  
DB 5108 GCGCGCGCGCGCGAGAGCGCTGACGCTGATCTCTCTTCGCGCGCGCGGACGCAAG 5167  
QY 1122 GCGCGCGATTTCCATACCTGCGGAGCGGTGCGCGCTTCGCGCGGCGGTGCTGACTG 1181  
DB 5168 GCGCGCGAGAACAGCGCTATTTCAAACCTTCGAGCGAAGCGCGCGGAGGCTGTC 5227  
QY 1182 CTTGGCGGTGACCGCGGCTGATTGCGCAGGCACTGGGCAACGCGGTACCGCTGCTGCGC 1241  
DB 5228 GCGCGCGAATACACGCGGAAATGGCCAGGTGAAGGCCATGCTG-----ACGTGCGC 5281  
QY 1242 GTCGCAACGCTGACGAAGAGTTCGCGGAGCGCGCGAGCTGCGCGCGGAGCGATGCG 1301

Db 5282 AAGGTCAGCTGGGACGTGTGAGTGGAGAGCCCGAACTGCTCGCGGCTCGGACGAG 5341  
QY 1302 GTGCTGTT 1309  
Db 5342 GGGCTGTT 5349

RESULT 14  
US-08-920-827-13  
; Sequence 13, Application US/08920827  
; Patent No. 5770375  
; GENERAL INFORMATION:  
; APPLICANT: Ohno, Tsuneya  
; APPLICANT: Matsuhisa, Akio  
; APPLICANT: Uehara, Hirotugu  
; APPLICANT: Eda, Soji  
; TITLE OF INVENTION: Probe for Diagnosing Infectious Disease  
; NUMBER OF SEQUENCES: 25  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
; STREET: 6300 Sears Tower, 233 South Wacker Drive  
; CITY: Chicago  
; STATE: Illinois  
; COUNTRY: United States of America  
; ZIP: 60606-8402  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patencin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/920,827  
; FILING DATE: 29-AUG-1997  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/362,577  
; FILING DATE: 27-MAR-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Rin-Laures, Li-Hsien  
; REGISTRATION NUMBER: 33,547  
; REFERENCE/DOCKET NUMBER: 19036/32420  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 312/474-6300  
; TELEFAX: 312/474-0448  
; TELEX: 25-3856  
; INFORMATION FOR SEQ ID NO: 13:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 9515 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: Genomic DNA  
; ORIGINAL SOURCE:  
; ORGANISM: Pseudomonas aeruginosa  
; STRAIN: Clinical Isolate P2-2  
; US-08-920-827-13

Query Match 5.2%; Score 76; DB 1; Length 9515;  
Best Local Similarity 44.3%; Pred. No. 4e-06;  
Matches 455; Conservative 0; Mismatches 560; Indels 13; Gaps 3;

QY 283 GCTTGTGCTGCGCACCCTGCTGTATACAGGCCCGCGAAGGCGTGCATCCG 342  
DB 4334 GCTTCCAGCGGACACCTTGTCCATCGCAAGGCGCTGAGCTCGGCTACATCCCATG 4393  
QY 343 GTGACATCATCTCTTCGCGCGGAGGGAAGGCCCGGATCGTCGCCATCACCGGTTCCA 402  
DB 4394 GCGGCTGTACTTCGCAAGCGCATCGCGAGGTGTGTGGAGAGGCGGGGTTCG 4453  
QY 403 ACCGGAAGAGCAGCACCTGTGTGGCGAAATGCGGTTGGCGGACAAACGCTG 462  
DB 4454 CCACGGCTGACCTATTCCGCGCACCCGCGGCGGCTGGCCATCGCCAACTCA 4513

[illegible]

RESULT 15

```

; RESOLUTION 13
; US-08-921-177-13
; Sequence 13, Application US/08921177
; Patent No. 5798211
; GENERAL INFORMATION:
; APPLICANT: Ohno, Tsuneya
; APPLICANT: Matsuhisa, Akio
; APPLICANT: Uehara, Hirotetsugu
; APPLICANT: Eda, Soji
; TITLE OF INVENTION: Probe for Diagnosing Infectious Disease
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
;

```



Search completed: August 14, 2003, 04:14:48  
Job time : 134 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 14, 2003, 04:40:26 ; Search time 420 Seconds  
(without alignments)  
7726.222 Million cell updates/sec

Title: US-09-701-229-1  
Perfect score: 1450  
Sequence: 1 cgtgctgctgctgctgcca.....tgttgacgcgcacgcgcgc 1450

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 1504479 seqs, 1118970152 residues

Word size : 0

Total number of hits satisfying chosen parameters: 3008958

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Published\_Applications\_NA.\*

- 1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq.\*
- 2: /cgn2\_6/ptodata/1/pubpna/PCT\_NEW\_PUB.seq.\*
- 3: /cgn2\_6/ptodata/1/pubpna/US06\_NEW\_PUB.seq.\*
- 4: /cgn2\_6/ptodata/1/pubpna/US06\_PUBCOMB.seq.\*
- 5: /cgn2\_6/ptodata/1/pubpna/US07\_NEW\_PUB.seq.\*
- 6: /cgn2\_6/ptodata/1/pubpna/PCTUS\_PUBCOMB.seq.\*
- 7: /cgn2\_6/ptodata/1/pubpna/US08\_NEW\_PUB.seq.\*
- 8: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq.\*
- 9: /cgn2\_6/ptodata/1/pubpna/US09A\_PUBCOMB.seq.\*
- 10: /cgn2\_6/ptodata/1/pubpna/US09B\_PUBCOMB.seq.\*
- 11: /cgn2\_6/ptodata/1/pubpna/US09C\_PUBCOMB.seq.\*
- 12: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq.\*
- 13: /cgn2\_6/ptodata/1/pubpna/US10A\_PUBCOMB.seq.\*
- 14: /cgn2\_6/ptodata/1/pubpna/US10B\_PUBCOMB.seq.\*
- 15: /cgn2\_6/ptodata/1/pubpna/US10\_NEW\_PUB.seq.\*
- 16: /cgn2\_6/ptodata/1/pubpna/US60\_NEW\_PUB.seq.\*
- 17: /cgn2\_6/ptodata/1/pubpna/US60\_PUBCOMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1347	92.9	1347	9	US-09-815-242-7912
2	54	3.7	1200	9	US-09-815-242-4132
3	36	2.5	116	9	US-09-815-242-1313
4	36	2.5	1083	9	US-09-815-242-7913
5	24	1.7	1317	9	US-09-741-669-259
6	24	1.7	1317	9	US-09-815-242-5932
7	20	1.4	1425	14	US-10-156-761-5271
8	20	1.4	2730	11	US-09-927-827-5
9	20	1.4	3511	10	US-09-753-831-3
10	20	1.4	7356	11	US-09-927-827-1
11	20	1.4	9025608	14	US-10-156-761-1
12	19	1.3	393	14	US-10-156-761-3485
13	19	1.3	522	12	US-10-084-843-268
14	19	1.3	522	12	US-10-193-002-263
15	19	1.3	648	14	US-10-156-761-7015
16	19	1.3	927	14	US-10-156-761-3786

17	1.3	954	14	US-10-156-761-31	Sequence 31, Appli
18	1.3	966	9	US-09-815-242-9789	Sequence 9789, Ap
19	1.3	2169	13	US-10-098-841-266	Sequence 266, App
20	1.3	2932	14	US-10-225-5678-447	Sequence 447, App
21	1.3	3489	9	US-09-815-242-7731	Sequence 7731, Ap
22	1.3	4725	14	US-10-205-032-17	Sequence 17, Appl
23	1.3	11238	14	US-10-205-032-15	Sequence 15, Appl
24	1.3	14272	9	US-09-870-162A-23	Sequence 23, Appl
25	1.3	60196	14	US-10-205-032-1	Sequence 1, Appli
26	1.3	9025608	14	US-10-156-761-1	Sequence 1, Appli
27	1.2	938	13	US-10-027-632-162057	Sequence 162057,
28	1.2	1017	14	US-10-156-761-1816	Sequence 1816, Ap
29	1.2	1446	9	US-09-815-242-7925	Sequence 7925, Ap
30	1.2	1626	14	US-10-156-761-4834	Sequence 4834, Ap
31	1.2	1719	14	US-10-156-761-2779	Sequence 2779, Ap
32	1.2	1883	13	US-10-033-109-9	Sequence 9, Appli
33	1.2	1941	14	US-10-156-761-2185	Sequence 2185, Ap
34	1.2	2205	14	US-10-156-761-4155	Sequence 4155, Ap
35	1.2	2271	12	US-10-101-510-133	Sequence 133, App
36	1.2	2487	14	US-10-205-823-135	Sequence 135, App
37	1.2	2712	14	US-10-156-761-1212	Sequence 1212, Ap
38	1.2	4041	10	US-09-861-289-36	Sequence 36, Appl
39	1.2	4041	10	US-09-860-846-36	Sequence 36, Appl
40	1.2	4041	11	US-09-988-384B-36	Sequence 36, Appl
41	1.2	4041	11	US-09-836-821-36	Sequence 36, Appl
42	1.2	36778	10	US-09-861-289-5	Sequence 5, Appli
43	1.2	36778	10	US-09-860-846-5	Sequence 5, Appli
44	1.2	36778	11	US-09-836-821-5	Sequence 5, Appli
45	1.2	37948	11	US-09-988-384B-5	Sequence 5, Appli

ALIGNMENTS

RESULT 1  
US-09-815-242-7912  
Sequence 7912, Application US/09815242  
Patent No. US20020061569A1

GENERAL INFORMATION:  
APPLICANT: Haselbeck, Robert  
APPLICANT: Ohlsen, Kari L.  
APPLICANT: Zyskind, Judith W.  
APPLICANT: Wall, Daniel  
APPLICANT: Trawick, John D.  
APPLICANT: Carr, Grant J.  
APPLICANT: Yamamoto, Robert T.  
APPLICANT: Xu, H. Howard  
TITLE OF INVENTION: Identification of Essential Genes in  
TITLE OF INVENTION: Prokaryotes

FILE REFERENCE: ELITRA.011A  
CURRENT APPLICATION NUMBER: US/09/815,242  
CURRENT FILING DATE: 2001-03-21  
PRIOR APPLICATION NUMBER: 60/191,078  
PRIOR FILING DATE: 2000-03-21  
PRIOR APPLICATION NUMBER: 60/206,848  
PRIOR FILING DATE: 2000-05-23  
PRIOR APPLICATION NUMBER: 60/207,727  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: 60/242,578  
PRIOR FILING DATE: 2000-10-23  
PRIOR APPLICATION NUMBER: 60/253,625  
PRIOR FILING DATE: 2000-11-27  
PRIOR APPLICATION NUMBER: 60/257,931  
PRIOR FILING DATE: 2000-12-22  
PRIOR APPLICATION NUMBER: 60/269,308  
PRIOR FILING DATE: 2001-02-16

NUMBER OF SEQ ID NOS: 14110  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 7912  
LENGTH: 1347  
TYPE: DNA  
ORGANISM: Pseudomonas aeruginosa  
FEATURE:

Handwritten signature: *W. J. ...*

; NAME/KEY: CDS			
; LOCATION: (1)...(1347)			
US-09-815-242-7912			
Query Match		92.9%; Score 1347; DB 9; Length 1347;	
Best Local Similarity		100.0%; Pred. No. 0;	
Matches 1347; Conservative		0; Mismatches 0; Indels 0; Gaps 0;	
QY	51	ATGAGCCTGATCGCCTCCGACCACTTCGCGATCGTTGTGCGCCCTCGGCAAGAGCGCATG	110
DB	1	ATGAGCCTGATCGCCTCCGACCACTTCGCGATCGTTGTGCGCCCTCGGCAAGAGCGCATG	60
QY	111	TCCTGGTGGGTACCTGGCGGCGCGGCTTGCCTTTCCCGGTGTCGATACCCGAGAG	170
DB	61	TCCTGGTGGGTACCTGGCGGCGCGGCTTGCCTTTCCCGGTGTCGATACCCGAGAG	120
QY	171	AACTCCCGGAGTGGCCACCTTCGCTGCGCGCGGCTTGCCTTTCCCGGTGTCGATACCCGAGAG	230
DB	121	AACTCCCGGAGTGGCCACCTTCGCTGCGCGCGGCTTGCCTTTCCCGGTGTCGATACCCGAGAG	180
QY	231	GAATCGAGCGCGAGTTCCTGCTCGCGCGGAACTTATGTCTAGCGCCGCGCTTGTGCG	290
DB	181	GAATCGAGCGCGAGTTCCTGCTCGCGCGGAACTTATGTCTAGCGCCGCGCTTGTGCG	240
QY	291	CTGCGCACCCCTGCGTGTGTACAGCGCGCGGAAAGCGTGGCATCTCCGCTGACATC	350
DB	241	CTGCGCACCCCTGCGTGTGTACAGCGCGCGGAAAGCGTGGCATCTCCGCTGACATC	300
QY	351	GATCTCTTCCCGCGAGCGGAGGCGCGCGGATCGTGCCTATCACCGTTCCAGCGGAAG	410
DB	301	GATCTCTTCCCGCGAGCGGAGGCGCGCGGATCGTGCCTATCACCGTTCCAGCGGAAG	360
QY	411	AGCACCGTGACCACTCGTGGCGAAATGGCGTGGCGCGGACAAAGCGTGTGCGCGTC	470
DB	361	AGCACCGTGACCACTCGTGGCGAAATGGCGTGGCGCGGACAAAGCGTGTGCGCGTC	420
QY	471	GGCGGCAACTCGGCACTCGGCGTGTGCGCGTGTGCGCGGACGACATCGAGTGTACGTC	530
DB	421	GGCGGCAACTCGGCACTCGGCGTGTGCGCGTGTGCGCGGACGACATCGAGTGTACGTC	480
QY	531	TTGGAGCTCTGAGCTTCAGCTGGAACTCGGATCGCTCAACCGCGAGTGGCGACC	590
DB	481	TTGGAGCTCTGAGCTTCAGCTGGAACTCGGATCGCTCAACCGCGAGTGGCGACC	540
QY	591	GTGCTGAAGCTCAGGCAACCACTATGGATCGCTACGCGGATGGCTCACTACCACTG	650
DB	541	GTGCTGAAGCTCAGGCAACCACTATGGATCGCTACGCGGATGGCTCACTACCACTG	600
QY	651	GCCAGCACCGATCTTCGCGGTCGCCCGGAGTGGTGGTGAATCGCGCGGATGCCCTG	710
DB	601	GCCAGCACCGATCTTCGCGGTCGCCCGGAGTGGTGGTGAATCGCGCGGATGCCCTG	660
QY	711	ACCCGACCGCTGATCGCCGATACCGTGGCGTGTGCTGCTGCTGCTGCTGCTGCTGCTG	770
DB	661	ACCCGACCGCTGATCGCCGATACCGTGGCGTGTGCTGCTGCTGCTGCTGCTGCTGCTG	720
QY	771	TTCAAGGCTTTTCGGCTGATCAGGAAAGCGGCGGAGTGGCTGGCGTTCCAGTTCCGAC	830
DB	721	TTCAAGGCTTTTCGGCTGATCAGGAAAGCGGCGGAGTGGCTGGCGTTCCAGTTCCGAC	780
QY	831	AACTGCTGCGGTTGGCGAACTGAAGTCCCTGGCGGCGGCACTATTTCCAAACCGCTC	890
DB	781	AACTGCTGCGGTTGGCGAACTGAAGTCCCTGGCGGCGGCACTATTTCCAAACCGCTC	840
QY	891	GCGCGCTGCGCTGGCGCATCGCTGGCGCTGGCGTTCGAGCGCATGCTCGCGCGCTG	950
DB	841	GCGCGCTGCGCTGGCGCATCGCTGGCGCTGGCGTTCGAGCGCATGCTCGCGCGCTG	900
QY	951	AAGCGCTTTTCGGCTGATCGCTGCGGAGTGGTACGCGAGCGGCGGCTGAGC	1010
DB	901	AAGCGCTTTTCGGCTGATCGCTGCGGAGTGGTACGCGAGCGGCGGCTGAGC	960
QY	1011	TACTACGAGGATTCCAAAGCCCAACAACTGTCGCGCGCGCTTGGCGGCGATCGAGGGGCTG	1070

DB	961	TACTACGAGGATTCCAAAGCCCAACAACTGTCGCGCGCGCTTGGCGGCGATCGAGGGGCTG	1020
QY	1071	GGTGGCGACATCGAGCGCAAGCTGGTGTGCTGCGCGCGGAGAGCGGCGCGCAT	1130
DB	1021	GGTGGCGACATCGAGCGCAAGCTGGTGTGCTGCGCGCGGAGAGCGGCGCGCAT	1080
QY	1131	TTCCATGACCTGCGCGAGCGGCTGCGCGCTTCTGCGCGGCGGCTGCTGCTGCGCGCT	1190
DB	1081	TTCCATGACCTGCGCGAGCGGCTGCGCGCTTCTGCGCGGCGGCTGCTGCTGCGCGCT	1140
QY	1191	GACCGCGGCTGATTGCCCGAGGACTGGGCAACGCGGTACCGCTGTGCGCGTGGCAACG	1250
DB	1141	GACCGCGGCTGATTGCCCGAGGACTGGGCAACGCGGTACCGCTGTGCGCGTGGCAACG	1200
QY	1251	CTGGAGGAGGAGTCCGCGAGCGCGCGGAGCTGGCGGAGGCGATCGGCTGCTGTTG	1310
DB	1201	CTGGAGGAGGAGTCCGCGAGCGCGCGGAGCTGGCGGAGGCGATCGGCTGCTGTTG	1260
QY	1311	TGCGCGGCTGCGCGAGCTGGACATGTTCAAGAACTTCGAAGAACGCGGAGCGCTGTTT	1370
DB	1261	TGCGCGGCTGCGCGAGCTGGACATGTTCAAGAACTTCGAAGAACGCGGAGCGCTGTTT	1320
QY	1371	GCCAAAGCGGTAGAGGAGTACGCTGA	1397
DB	1321	GCCAAAGCGGTAGAGGAGTACGCTGA	1347

RESULT 2

US-09-815-242-4132

; Sequence 4132, Application US/09815242

; Patent No. US20020061569A1

GENERAL INFORMATION:

; APPLICANT: Haselbeck, Robert

; APPLICANT: Ohlsen, Karl L.

; APPLICANT: Zyskind, Judith W.

; APPLICANT: Wall, Daniel

; APPLICANT: Trawick, John D.

; APPLICANT: Cart, Grant J.

; APPLICANT: Yamamoto, Robert T.

; APPLICANT: Xu, H. Howard

; TITLE OF INVENTION: Identification of Essential Genes in

; FILE REFERENCE: ELITRA.011a

; CURRENT FILING DATE: 2001-03-21

; PRIOR APPLICATION NUMBER: 60/191,078

; PRIOR FILING DATE: 2000-03-21

; PRIOR APPLICATION NUMBER: 60/206,848

; PRIOR FILING DATE: 2000-05-23

; PRIOR APPLICATION NUMBER: 60/207,727

; PRIOR FILING DATE: 2000-05-26

; PRIOR APPLICATION NUMBER: 60/242,578

; PRIOR FILING DATE: 2000-10-23

; PRIOR APPLICATION NUMBER: 60/253,625

; PRIOR FILING DATE: 2000-11-27

; PRIOR APPLICATION NUMBER: 60/257,931

; PRIOR FILING DATE: 2000-12-22

; PRIOR APPLICATION NUMBER: 60/269,308

; PRIOR FILING DATE: 2001-02-16

; NUMBER OF SEQ ID NOS: 14110

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 4132

; LENGTH: 1200

; TYPE: DNA

; ORGANISM: Pseudomonas aeruginosa

US-09-815-242-4132

Query Match

Best Local Similarity

Matches 54; Conservative

3.7%; Score 54; DB 9; Length 1200;

100.0%; Pred. No. 1.9e-17;

0; Mismatches 0; Indels 0; Gaps 0;

QY 1397 | ATGCTGTCGGTGTTCGCGCCCTTCCCGTCCGCGCTGTTGAGCGCGGCGCATC | 1450 |

Db 1 ATGCTGTCGGTGTTCGCCCTTCCGRCGCGCTGTGTGAGCGGCACGGCATC 54

RESULT 3

US-09-815-242-1313/c  
; Sequence 1313, Application US/09815242  
; Patent No. US20020061569A1

GENERAL INFORMATION:

; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Kari L.  
; APPLICANT: Zyskind, Judith W.  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John D.  
; APPLICANT: Carr, Grant J.  
; APPLICANT: Yamamoto, Robert T.  
; APPLICANT: Xu, H. Howard

; TITLE OF INVENTION: Identification of Essential Genes in

; TITLE OF INVENTION: Prokaryotes

; FILE REFERENCE: ELITRA.011A

; CURRENT APPLICATION NUMBER: US/09/815,242

; PRIOR FILING DATE: 2001-03-21

; PRIOR APPLICATION NUMBER: 60/191,078

; PRIOR FILING DATE: 2000-03-21

; PRIOR APPLICATION NUMBER: 60/206,848

; PRIOR FILING DATE: 2000-05-23

; PRIOR APPLICATION NUMBER: 60/207,727

; PRIOR FILING DATE: 2000-05-26

; PRIOR APPLICATION NUMBER: 60/242,578

; PRIOR FILING DATE: 2000-10-23

; PRIOR APPLICATION NUMBER: 60/253,625

; PRIOR FILING DATE: 2000-11-27

; PRIOR APPLICATION NUMBER: 60/257,931

; PRIOR FILING DATE: 2000-12-22

; PRIOR APPLICATION NUMBER: 60/269,308

; PRIOR FILING DATE: 2001-02-16

; NUMBER OF SEQ ID NOS: 14110

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 1313

; LENGTH: 116

; TYPE: DNA

; ORGANISM: Pseudomonas aeruginosa

US-09-815-242-1313

Query Match 2.5%; Score 36; DB 9; Length 116;

Best Local Similarity 100.0%; Pred. No. 4.1e-08;

Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1415 CCCTTCCCGTCGCGCTTTGAGCGGCACGGCATC 1450

Db 116 CCCTTCCCGTCGCGCTTTGAGCGGCACGGCATC 81

RESULT 4

US-09-815-242-7913

; Sequence 7913, Application US/09815242  
; Patent No. US20020061569A1

GENERAL INFORMATION:

; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Kari L.  
; APPLICANT: Zyskind, Judith W.  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John D.  
; APPLICANT: Carr, Grant J.  
; APPLICANT: Yamamoto, Robert T.  
; APPLICANT: Xu, H. Howard

; TITLE OF INVENTION: Identification of Essential Genes in

; TITLE OF INVENTION: Prokaryotes

; FILE REFERENCE: ELITRA.011A

; CURRENT APPLICATION NUMBER: US/09/815,242

; PRIOR FILING DATE: 2001-03-21

; PRIOR APPLICATION NUMBER: 60/191,078

; PRIOR FILING DATE: 2000-03-21

; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/269,308  
; PRIOR FILING DATE: 2001-02-16  
; NUMBER OF SEQ ID NOS: 14110  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 7913

; LENGTH: 1083

; TYPE: DNA

; ORGANISM: Pseudomonas aeruginosa

; FEATURE:

; NAME/KEY: CDS

; LOCATION: (1)...(1083)

US-09-815-242-7913

Query Match 2.5%; Score 36; DB 9; Length 1083;

Best Local Similarity 100.0%; Pred. No. 3.1e-08;

Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GTGCTGATCGCGCTCGCCACCTTGAAGCTGCGTTGA 37

Db 1048 GTGCTGATCGCGCTCGCCACCTTGAAGCTGCGTTGA 1083

RESULT 5

US-09-741-669-259

; Sequence 259, Application US/09741669  
; Patent No. US20020022718A1

GENERAL INFORMATION:

; APPLICANT: Forsyth, R. Allyn  
; APPLICANT: Ohlsen, Kari L.  
; APPLICANT: Zyskind, Judith W.  
; TITLE OF INVENTION: Genes identified as required for  
; TITLE OF INVENTION: proliferation of E. coli  
; FILE REFERENCE: ELITRA.009A

; CURRENT APPLICATION NUMBER: US/09/741,669

; CURRENT FILING DATE: 2000-12-19

; PRIOR APPLICATION NUMBER: US 60/173005

; PRIOR FILING DATE: 1999-12-23

; NUMBER OF SEQ ID NOS: 481

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 259

; LENGTH: 1317

; TYPE: DNA

; ORGANISM: Escherichia coli

; FEATURE:

; NAME/KEY: CDS

; LOCATION: (1)...(1317)

US-09-741-669-259

Query Match 1.7%; Score 24; DB 9; Length 1317;

Best Local Similarity 100.0%; Pred. No. 0.041;

Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 537 CTGTCGAGCTTCCAGCTGGAACCC 560

Db 475 CTGTCGAGCTTCCAGCTGGAACCC 498

RESULT 6

US-09-815-242-5932

; Sequence 5932, Application US/09815242  
; Patent No. US20020061569A1

GENERAL INFORMATION:

; APPLICANT: Haselbeck, Robert

25

Db 2507 GCAGCGCGCGAGCTGGCGC 2488

## RESULT 10

US-09-927-827-1/c  
; Sequence 1, Application US/09927827  
; Publication No. US20030036176A1  
; GENERAL INFORMATION:  
; APPLICANT: Bower, Stanley G.  
; APPLICANT: Ramseier, Thomas M.  
; TITLE OF INVENTION: Directed Genetic Engineering of Xanthomonas campestris  
; FILE REFERENCE: 38-10(15824)B  
; CURRENT APPLICATION NUMBER: US/09/927,827  
; CURRENT FILING DATE: 2001-08-10  
; PRIOR APPLICATION NUMBER: US 60/279,493  
; PRIOR FILING DATE: 2001-03-28  
; NUMBER OF SEQ ID NOS: 69  
; SEQ ID NO 1  
; LENGTH: 7356  
; TYPE: DNA  
; ORGANISM: Xanthomonas campestris  
; FEATURE:  
; NAME/KEY: unsure  
; LOCATION: (1)..(7356)  
; OTHER INFORMATION: unsure at all n locations  
US-09-927-827-1

Query Match 1.4%; Score 20; DB 11; Length 7356;

Best Local Similarity 100.0%; Pred. No. 3.6;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 500 CCGTGGCGCGAGCATCG 519

|||||

Db 6090 CCGTGGCGCGAGCATCG 6071

## RESULT 11

US-10-156-761-1/c  
; Sequence 1, Application US/10156761  
; Publication No. US20030119018A1  
; GENERAL INFORMATION:  
; APPLICANT: OMURA, SATOSHI  
; APPLICANT: IKEDA, HARUO  
; APPLICANT: ISHIKAWA, JUN  
; APPLICANT: HIRAKAWA, HIROSHI  
; APPLICANT: SHIBA, TADAYOSHI  
; APPLICANT: SAKAKI, YOSHIYUKI  
; APPLICANT: HATTORI, MASAHIRA  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-262  
; CURRENT APPLICATION NUMBER: US/10/156,761  
; CURRENT FILING DATE: 2002-05-29  
; PRIOR APPLICATION NUMBER: JP 2001-204089  
; PRIOR FILING DATE: 2001-05-30  
; PRIOR APPLICATION NUMBER: JP 2001-272697  
; PRIOR FILING DATE: 2001-08-02  
; NUMBER OF SEQ ID NOS: 15109  
; SEQ ID NO 1  
; LENGTH: 9025608  
; TYPE: DNA  
; ORGANISM: Streptomyces avermitilis  
; FEATURE:  
; NAME/KEY: misc.feature  
; LOCATION: (4187715)  
; OTHER INFORMATION: a, t, c, g, other or unknown  
US-10-156-761-1

Query Match 1.4%; Score 20; DB 14; Length 9025608;

Best Local Similarity 100.0%; Pred. No. 1.4;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1094 GGTGCTGCTCGCGCGGAG 1113

|||||

Db 6407532 GGTGCTGCTCGCGCGGAG 6407513

## RESULT 12

US-10-156-761-3485  
; Sequence 3485, Application US/10156761  
; Publication No. US20030119018A1  
; GENERAL INFORMATION:  
; APPLICANT: OMURA, SATOSHI  
; APPLICANT: IKEDA, HARUO  
; APPLICANT: ISHIKAWA, JUN  
; APPLICANT: HIRAKAWA, HIROSHI  
; APPLICANT: SHIBA, TADAYOSHI  
; APPLICANT: SAKAKI, YOSHIYUKI  
; APPLICANT: HATTORI, MASAHIRA  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-262  
; CURRENT APPLICATION NUMBER: US/10/156,761  
; CURRENT FILING DATE: 2002-05-29  
; PRIOR APPLICATION NUMBER: JP 2001-204089  
; PRIOR FILING DATE: 2001-05-30  
; PRIOR APPLICATION NUMBER: JP 2001-272697  
; PRIOR FILING DATE: 2001-08-02  
; NUMBER OF SEQ ID NOS: 15109  
; SEQ ID NO 3485  
; LENGTH: 393  
; TYPE: DNA  
; ORGANISM: Streptomyces avermitilis  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)..(393)  
US-10-156-761-3485

## Query Match

1.3%; Score 19; DB 14; Length 393;  
Best Local Similarity 100.0%; Pred. No. 17;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1270 AGCGCGCGAGCTGGCGC 1288

|||||

Db 143 AGCGCGCGAGCTGGCGC 161

## RESULT 13

US-10-084-843-268  
; Sequence 268, Application US/10084843  
; Publication No. US20030143243A1  
; GENERAL INFORMATION:  
; APPLICANT: Reed, Steven G.  
; APPLICANT: Skeiky, Yasir A.W.  
; APPLICANT: Dillon, Davin C.  
; APPLICANT: Campos-Neto, Antonio  
; APPLICANT: Houghton, Raymond  
; APPLICANT: Vedvick, Thomas S.  
; APPLICANT: Twardzik, Daniel R.  
; APPLICANT: Lodes, Michael J.  
; APPLICANT: Hendrickson, Ronald C.  
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY  
; AND DIAGNOSIS OF TUBERCULOSIS  
; NUMBER OF SEQUENCES: 355  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: SEED and BERRY LLP  
; STREET: 6300 Columbia Center, 701 Fifth Avenue  
; CITY: Seattle  
; STATE: Washington  
; COUNTRY: USA  
; ZIP: 98104-7092  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/10/084,843

```
;/ FILING DATE: 25-Feb-2002
;/ CLASSIFICATION: <Unknown>
;/ PRIOR APPLICATION DATA:
;/ APPLICATION NUMBER: US/09/072,967
;/ FILING DATE: 05-MAY-1998
;/ ATTORNEY/AGENT INFORMATION:
;/ NAME: Makl, David J.
;/ REGISTRATION NUMBER: 31,392
;/ TELECOMMUNICATION INFORMATION:
;/ TELEPHONE: (206) 622-4900
;/ TELEFAX: (206) 682-6031
;/ INFORMATION FOR SEQ ID NO: 268:
;/ SEQUENCE CHARACTERISTICS:
;/ LENGTH: 522 base pairs
;/ TYPE: nucleic acid
;/ STRANDEDNESS: single
;/ TOPOLOGY: linear
;/ MOLECULE TYPE: cDNA
;/ SEQUENCE DESCRIPTION: SEQ ID NO: 268:
US-10-084-843-268

Query Match 1.3%; Score 19; DB 12; Length 522;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1082 CGACGGCAAGCTGGTCTG 1100
Db 429 CGACGGCAAGCTGGTCTG 447

RESULT 14
US-10-193-002-263
; Sequence 263, Application US/10193002
; Publication No. US20030135026A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; Skelky, Yasir A.W.
; Dillon, Davin C.
; Campos-Neto, Antonia
; Houghton, Raymond
; Vedvick, Thomas S.
; Twardzik, Daniel R.
; Lodes, Michael J.
; Hendrickson, Ronald C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF
; TUBERCULOSIS
; NUMBER OF SEQUENCES: 350
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/193,002
; FILING DATE: 10-Jul-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/072,596
; FILING DATE: 05-MAY-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Makl, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.417C9
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
```

```
;/ TELEFAX: (206) 682-6031
;/ INFORMATION FOR SEQ ID NO: 263:
;/ SEQUENCE CHARACTERISTICS:
;/ LENGTH: 522 base pairs
;/ TYPE: nucleic acid
;/ STRANDEDNESS: single
;/ TOPOLOGY: linear
;/ MOLECULE TYPE: cDNA
;/ SEQUENCE DESCRIPTION: SEQ ID NO: 263:
US-10-193-002-263

Query Match 1.3%; Score 19; DB 12; Length 522;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1082 CGACGGCAAGCTGGTCTG 1100
Db 429 CGACGGCAAGCTGGTCTG 447

RESULT 15
US-10-156-761-7015
; Sequence 7015, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 7015
; LENGTH: 648
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(648)
; US-10-156-761-7015

Query Match 1.3%; Score 19; DB 14; Length 648;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CGTGCTGATCGGCTCGCC 19
Db 546 CGTGCTGATCGGCTCGCC 564

Search completed: August 14, 2003, 07:13:43
Job time : 437 secs
```



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run On: August 14, 2003, 08:41:46 ; Search time 4238 seconds  
(without alignments)  
4324.566 Million cell updates/sec

Title: US-09-701-229-2

Perfect score: 448

Sequence: 1 MSLIASDHFRIVVGLKSGM.....MFKNFEERGLFAKAVEELA 448

Scoring table: OLIGO

Xgapop 60.0, Xgapext 60.0  
Ygapop 60.0, Ygapext 60.0  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 2888711 seqs, 2045481386 residues

Word size: 1

Total number of hits satisfying chosen parameters: 4990538

Minimum DB seq length: 25

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlp  
-O=/cgn2.1/USPTO.spool/US09701229/runat\_09082003\_161236\_5021/app.query.fasta.1.647  
-DB=GenEmbl -QFMT=fastap -SUFFIX=olip2n.rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=Dits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi -LIST=45  
-DOCCALIGN=200 -THR\_SCORE=quality -THR\_MIN=1 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc  
-NORM=ext -HEAPSIZ=500 -MINLEN=25 -MAXLEN=2000000000  
-USER=US09701229.CGN.1.1.5265 -runat\_09082003\_161236\_5021 -NCPU=3  
-NO\_MMAB -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database :

GenEmbl.\*

1: gb\_ba.\*  
2: gb\_hgt.\*  
3: gb\_in.\*  
4: gb\_om.\*  
5: gb\_ov.\*  
6: gb\_pat.\*  
7: gb\_ph.\*  
8: gb\_pl.\*  
9: gb\_pr.\*  
10: gb\_ro.\*  
11: gb\_sts.\*  
12: gb\_sy.\*  
13: gb\_un.\*  
14: gb\_vi.\*  
15: em\_ba.\*  
16: em\_fun.\*  
17: em\_hum.\*  
18: em\_in.\*  
19: em\_mu.\*  
20: em\_om.\*  
21: em\_or.\*  
22: em\_ov.\*  
23: em\_pat.\*  
24: em\_ph.\*  
25: em\_pl.\*  
26: em\_ro.\*  
27: em\_sts.\*  
28: em\_un.\*

29: em\_vi.\*  
30: em\_hgt\_hum.\*  
31: em\_hgt\_inv.\*  
32: em\_hgt\_other.\*  
33: em\_hgt\_mus.\*  
34: em\_hgt\_pln.\*  
35: em\_hgt\_rod.\*  
36: em\_hgt\_mam.\*  
37: em\_hgt\_vrt.\*  
38: em\_sy.\*  
39: em\_hgt\_hum.\*  
40: em\_hgt\_mus.\*  
41: em\_hgt\_other.\*

pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
c 1	448	100.0	24000	1	AE004856 Pseudomon
2	290	64.7	5280	1	AY008276 Pseudomon
3	35	7.8	301995	1	AE016779 Pseudomon
c 4	20	4.5	311600	1	AE016871 Pseudomon
c 5	19	4.2	96109	6	AX067460 Sequence
6	14	3.1	1317	6	AX189058 Sequence
7	14	3.1	1629	1	ECMURD
8	14	3.1	2608	1	ECMUROY
9	14	3.1	11441	1	AE015046 Shigella
10	14	3.1	11498	6	U32793 Haemophilus
11	14	3.1	12791	1	AX191762 Sequence
12	14	3.1	12757	1	AE005185 Escherich
13	14	3.1	21757	1	AE000118 Escherich
c 14	14	3.1	22201	1	AE004310 Vibrio ch
15	14	3.1	22348	1	AE008699 Salmonell
16	14	3.1	28277	6	AX191720 Sequence
17	14	3.1	28277	15	EC2MIN
18	14	3.1	110000	6	AR274513_11
19	14	3.1	110000	6	AR274513_12
20	14	3.1	111408	1	EC0110K
21	14	3.1	248308	2	AC073937 Mus muscu
22	14	3.1	251050	1	AL627265 Salmonell
c 23	14	3.1	259498	2	AC020876 Mus muscu
24	14	3.1	281530	1	AP002550 Escherich
25	14	3.1	290029	1	AE016978 Shigella
26	14	3.1	300169	1	AE016834 Salmonell
27	14	3.1	300409	1	AE016755 Escherich
c 28	14	3.1	301442	1	AE016798 Vibrio vu
29	13	2.9	597	1	AY142808 Helicobac
c 30	13	2.9	14643	1	AE014097 Buchnera
31	13	2.9	298900	1	AP005074 Nitrosomo
32	13	2.9	316050	1	EX321859 Nitrosomo
c 33	12	2.7	11726	1	AE013965 Yersinia
34	12	2.7	12232	1	AE006048 Pasteurel
c 35	12	2.7	13071	1	AE015855 Shewanell
36	12	2.7	21000	1	AB052554 Shewanell
c 37	12	2.7	188050	1	AL646072 Ralstonia
38	12	2.7	201050	1	AJ414143 Yersinia
39	12	2.7	301727	1	AE016960 Coxiella
c 40	12	2.7	348600	1	AE0163521
41	11	2.5	1497	1	EC0FTSW
42	11	2.5	110000	2	LMFLCHR34_08
c 43	11	2.5	166560	5	AL954741 Zebrafish
44	10	2.2	10629	1	AE002398 Neisseria
c 45	10	2.2	11683	1	AE013119 Thermosoma

# ALIGNMENTS

RESULT 1



```

previously reported genes of unknown function, or no
similarity to any previously reported sequences)"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAG07793.1"
/db_xref="GI:9950637"
/translation="MLFAAQRIQRIQLLESOLQAPAPYCHVASWGRGCLLLIVTD
GHWATRLRYQRLQRLQQLAQQFEFANLAKILFKVQPTTSTQRAARSLTISAVAENL
RSAAETVGNPRLREALERLASRAEKSQRE"
complement(5865..6776)
/genes="lpxC"
/locus_tag="PA4406"
/note="synonyms: envA, asmb"
complement(5865..6776)
/genes="lpxC"
/locus_tag="PA4406"
/note="Protein name confidence: Class 1 (function
experimentally demonstrated in P. aeruginosa); Alt protein
name: EnvA protein"
/codon_start=1
/transl_table=11
/product="UDP-3-O-acyl-N-acetylglucosamine deacetylase"
/protein_id="AAG07794.1"
/db_xref="GI:9950638"
/translation="MIKQRTLKNIRATGVLSHGSEKVIYTLKPAPVDGTGIVFCRTDL
DPVEIAPRAENVGETMTSLVKGVDVDTVEHLLSAMAGLIDNAYVELSASEYPI
MDGAPGFVFLQISAGLQEAQAKKIRIKREVSVVEEGDKRAVFPDFGFKVSEFID
HDSVFGRTQQASVDSSTSFVKEVSARATFGMRDIEYLRSONLALGGSVENAIIVD
ENRVLEDGLRVEDEFVKHILDAIGDLVLLGNSLIGEPFGKSGHALNNQLLRTLIA
DKDAEWVTFEDARTAPISVMRPAAV"
complement(6889..8073)
/genes="ftsZ"
/locus_tag="PA4407"
complement(6889..8073)
/genes="ftsZ"
/locus_tag="PA4407"
/note="Protein name confidence: Class 2 (high similarity
to functionally studied protein)"
/codon_start=1
/transl_table=11
/product="cell division protein FtsZ"
/protein_id="AAG07795.1"
/db_xref="GI:9950639"
/translation="MFELVDNIAQTAVIKIVYGGGGNNAVNHMAKNNVGVEFICAN
TDAQALNTAARTVLQIGPGVTGLGAGANPEVGRQAALREDRERISEVLEGADMVIT
TGMGGGTGAATPIAIAEVAKEMILTAVVYVTFPEGRKRMQIADEGIRALAESYDS
LITIPNEKLLTILGKDALLAAPKADDVLGAVRGISDIIKEPGMINVDYFADVKTM
SEMGAMNMGTCASGNPRAEATEAARINPLEDVLNQLGARGILVNIATGPDLSLEY
SDVGNIEGFASHEATVIVKGVTVIADMRDELHVTVVATGLGARLEKPKVKVDNVTGVS
AQAAPAPRAQREQSVNRYDLDRPTVMRNHSGSAATAAKLINPQDDILDYLDIPAFRLRR
AD"
complement(8124..9377)
/genes="ftsA"
/locus_tag="PA4408"

gene

Alignment Scores:
Pred. No.: 0 Length: 24000
Score: 448.00 Matches: 448
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 1 Gaps: 0

US-09-701-229-2 (1-448) x AE004856 (1-24000)

Qy 1 MetSerLeuIlealaSerAspHisPheArgIleValValcglyLeuGlytySerGlyMet 20
Db 16265 ATGAGCTGATGCCCTCCGACCACCTTCGCGCATCGTTGTCGCGCTCGCAAGACGCGCATG 16200

Qy 21 SerLeuValArgTyrLeuAlaArgArgGlyLeuProPheAlaValValAspThrArgGlu 40
Db 16205 TCCTGCTGGTCTACTTGGCGCGCGCGGTTCCTTTTCGCTGGTTCGATACCCGAGAG 16140

```

```
QY 401 LeuAspGluAlaValArgGlnAlaAlaGluLeuAlaArgGluGlyAspAlaValLeuLeu 420
Db 15065 CTTGAGCAGACAGCTCCGCGAGCGCGAGCTGCGCGGCGGAGCGATGCGGTCTGCTTT 15006
QY 421 SerProAlaCysAlaSerLeuAspMetPheLysAsnPheGluGluArgGlyArgLeuPhe 440
Db 15005 TCCCGCGCGCTCCGCGAGCGCTGACATGTTCAAGAACTTCAAGAACCGCGGACCGCTGCTC 14946
QY 441 AlaLysAlaValGluGluLeuAla 448
Db 14945 GCCAAGCGCTAGAGGAGCTAGCG 14922

RESULT 2
AY008276
LOCUS
DEFINITION
Pseudomonas aeruginosa MurE (murE), MurF (murF), MurG (murG), and
Murd (murD) genes, complete cds.
ACCESSION
AY008276
VERSION
AY008276.2
KEYWORDS
GI:12007715
SOURCE
Pseudomonas aeruginosa
ORGANISM
Pseudomonas aeruginosa
Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
Pseudomonadaceae; Pseudomonas.
1 (bases 1 to 5280)
Azzollina, B.A., Yuan, X., Anderson, M.S. and El-Sherbeini, M.
The cell wall and cell division gene cluster in the Mra operon of
Pseudomonas aeruginosa: cloning, production, and purification of
active enzymes
Protein Expr. Purif. 21 (3), 393-400 (2001)
21178826
11281713
2 (bases 1 to 5280)
El-Sherbeini, M. and Azzollina, B.
Pseudomonas aeruginosa murA, murD, murE, and murF genes
Unpublished
3 (bases 1 to 5280)
El-Sherbeini, M. and Azzollina, B.
Direct Submission
Submitted (20-SEP-2000) Biochemistry, Merck Research Laboratories,
126 E. Lincoln Avenue, Rahway, NJ 07065, USA
4 (bases 1 to 5280)
El-Sherbeini, M. and Azzollina, B.
Direct Submission
Submitted (02-JAN-2001) Biochemistry, Merck Research Laboratories,
126 E. Lincoln Avenue, Rahway, NJ 07065, USA
REMARK
Sequence update by submitter
COMMENT
On Jan 2, 2001 this sequence version replaced gi:10719753.
FEATURES
Location/Qualifiers
1..5280
/organism="Pseudomonas aeruginosa"
/mol_type="genomic DNA"
/db_xref="taxon:287"
1..1464
/gene="murE"
1..1464
/gene="murE"
/notes="UDP-N-acetyl-muramoyl-L-alanyl-D-glutamate:meso-2,
6-diaminopimelate ligase"
/codon_start=1
/transl_table=11
/product="MurE"
/protein_id="AAG45235.1"
/db_xref="GI:12007716"
/translation="MPSLSQLFQQAERDLIRELTLDHSGVRPVDLFTVPGHQDG
RAHIDALTAKGATVAYAEAGELPPSDAPLIAVKGLAAQLSAGVGRFYGPSPGLD
LIGVTGKTSVSOLVAQALDLIGERCISVCTGCTGFYGALESGRHTTDPPLAVQAT
LATLQACARAVAMEVSSHGLDQCRVALGEDIAVFTNLSRDLHYHGSMEYAAAKA
KLFPWDLRCRVINDDDFGRKAGEQDSILTYSLTSSAFLYCRASPASPATAR
RSSTAAKGLGLSLPGLRNLNLAAGLGLGYPGLDILRLTLPQGPVGRQRL
GGGKPLVVDYAHPTFDALKEVLRPHAAKRLICLFCGGDRDAGRKPLMAAIAER
LADEVLVDDNPRTASAAIADIRKFAAADKVTFLPSRGEAIAHLTASAAVDDVL
LAGKHEDYQIEDGVHRHPSDIEQAERALAAVEPHA"

QY 1 MetSerLeuIleAlaSerAspHisPheArgIleValValGlyLeuGlyLysSerGlyMet 20
|||||
```

```
1457..2833
/genes="murF"
1457..2833
/genes="murF"
/notes="recombinant enzyme catalyzed the ATP-dependent
addition of D-alanine-D-alanine to the
UDP-N-acetylmuramyl-L-alanine-D-glutamate-m-Dap precursor"
/codon_start=1
/transl_table=11
/product="MurF"
/protein_id="AAG45236.1"
/db_xref="GI:12007717"
/translation="MLEPLRLSQLTVALDARLIGEDAVFSAVTSDSRAIGPQGLFIAL
SGPRFDGHDYLAEEAAKGAVALVEREADAPLPOLLVRDTRALGLRGLNRRKFTG
PLAAMTSSGSKTAVKEMLASILRTQAGDAESVLATGNLNDIGVPLTLQLAPQHR
AVTELGASRIGETAVTETRHVAITNAGTAHVGEFGPEKIVKAEKLEGLAAD
GTAVNLDDKAFDTWKARASGRPLLTFSIDRFQADPRAADLQDRACGCGFRQGGVAG
EAQVNLNLGRHNVAALAAAHAALVPLDGIIVAGLQALQPKVAKRAVQLTAGKYSR
VIDDSYNANPASPAAIDILSGFSGRITVLVLDGMELGSMWAEQAHREVGAYAAKYSR
LYAVGPLMAHVAQFATGRHFAADQASLIGALATEQPTTTILIKGSRSAAMDKVAAL
CGSSESH"
2833..3915
/genes="mray"
2833..3915
/genes="mray"
/notes="phospho-N-acetylmuramoyl-pentapeptide translocase"
/codon_start=1
/transl_table=11
/product="Mray"
/protein_id="AAG22121.1"
/db_xref="GI:10719754"
/translation="MLLLAEYLQOQKFGVQYLTLRGILSVLTALSLSLWLGPM
IRTLAQIQGQAVRNDPQSHSKGTPMGALITAIATSLTLMWADSNRYVTVVL
VVTLLFGAIGWDDYDKVTEKNSRGLPSRWKVFQWSVFGIAGAVFLYMTAETPIETTL
IVPMLKSVELQIGIEFVLTLYFVIVGSSNAVALTDLGLAIMPVFMVAGALGIFCYL
SGNVPAEVLILPNVPGAGELIVFCALVAGLGLFWNTYPAQVFMVDVGDVALGAA
LGTIAVIVRQELVIFMGVFNETLSVMIOVASFKLTGRFRVFMAPIHHHFKGMP
DPRVIVRFWITVILVILGLATLKLK"
3929..5275
/genes="murD"
3929..5275
/genes="murD"
/notes="UDP-N-acetylmuramyl-L-alanine:D-glutamate ligase;
recombinant enzyme catalyzed the ATP-dependent
addition of D-glutamate to the precursor sugar peptide"
/codon_start=1
/transl_table=11
/product="MurD"
/protein_id="AAG45237.1"
/db_xref="GI:12007718"
/translation="MSLTASDHFRIVVGLGKSGMSLVRYLARGLPFVAVVDRENPE
LATLRAQYQVEARCGELDAEFLCSARELYVSPGLSLRTPALVQAAKGVRISSGIDL
FAEAKAPIAVITGNSNAKSTVTLVGMVAADKRVAVGGNLGTALDADDIELYV
LELSSFOLETCDRLNAEVATLVNVEDHMDRYDGMADYHLAKHRTFRGAROVVVRAD
ALTRPLIADTVPCWSEFGKPKDFKAGLIEEDGQKWLAFQFKLLPLVGGELKRGVNH
SNALAALALGHAVGLPFDMLGALKAFSLAHRCQWRERQKVSYYIDDKATNVGAAL
AAIEGLGADIDGKLVLAGGDKGADFHLREPVAFRCVAVLVLLRAGDLIAQALGNA
VPLRVATLDEAVRQAALAREGDAVLLSPACASLDMFKNFEEGRFLFAVEELA"
BASE COUNT 817 a 1781 c 1759 g 923 t
ORIGIN

Alignment Scores:
Pred. No.: 3,39e-291 Length: 5280
Score: 290.00 Matches: 446
Percent Similarity: 99.11% Conservative: 0
Best Local Similarity: 99.11% Mismatches: 2
Query Match: 64.73% Indels: 4
DB: 1 Gaps: 0

US-09-701-229-2 (1-448) x AY008276 (1-5280)
QY 1 MetSerLeuIleAlaSerAspHisPheArgIleValValGlyLeuGlyLysSerGlyMet 20
|||||
```

```

Db 3929 ATGAGCGTGATCCCTCCGACCACTTCCCGCATCGTTGTGGCGCTCCGGCAAGACGGCGATG 3988
QY 21 SerLeuValArgTyrLeuAlaArgArgGlyLeuProPheAlaValValAlaAspThrArgGlu 40
Db 3989 TCCCTGGTGGCTACCTTGGCGCGCGCGGCTTGCCTTTCCGCGTGCATACCCGAGAG 4048
QY 41 AsnProProGluLeuAlaThrLeuArgAlaGlnTyrProGlnValGluVal-ArgCysGI 60
Db 4049 AACCCCGCGAGTGGCTACCTTCGCTGCCGACATATCCGAGGTGGAAGC-GGTTTCGG 4107
QY 60 yGluLeuAspAlaGluPheLeuCysSerAlaArgGluLeuTyrValSerProGlyLeuSe 80
Db 4108 CGAACTCGACCGCGAGTTCCTTCGTCGCCCGCGAACTATATGACGCCCGCGCTGTC 4167
QY 80 rLeuArgThrProAlaLeuValGlnAlaAlaLysGlyValArgLysSerGlyAspII 100
Db 4168 GCTGCGCACCCCTGCGCTGGTACAGCGCGCGGAAAGGCGTGCATCTCCGCGTGCAT 4227
QY 100 eAspLeuPheAlaArgGluAlaLysAlaProIleValAlaIleThrGlySerAsnAlaLy 120
Db 4228 CGATCTCTTGGCCCGCGAGCGAAGCCCGCATCGTCGCATCACCAGGTTCACACGGA 4287
QY 120 sSerThrValThrLeuValGlyGluMetAlaValAlaAlaAspLysArgValAlaVa 140
Db 4288 GACACCGTGCACCCCTGTTGGCGGAATGGCGGTGGCGCGGACAAGCGTGTCCGCGT 4347
QY 140 lGlyLysAsnLeuGlyThrProAlaLeuAspLeuLeuAlaAlaAspIleGluLeuTyrVa 160
Db 4348 CGCGCGCAACCTCGCGCACCCCGCGCTGCACCTGCTGGCGGACGACATCGAGCTGTACGT 4407
QY 160 lLeuGluLeuSerSerPheGlnLeuGluThrCysAspArgLeuAsnAlaGluValAlaTh 180
Db 4408 GTTGGAGCTGTCGAGCTTCAGCTGGAAACCTCGCATCGCTCAACGCCGAGGTGGCGAC 4467
QY 180 rValLeuAsnValSerGluAspHisMetAspArgTyrAspGlyMetAlaAspTyrHisLe 200
Db 4468 CGTCTGAAGCTCAGCGAGACCATATGATCGCTACGAGCGGATGGCTGACTACCACT 4527
QY 200 uAlaLysHisArgIlePheArgGlyAlaArgGlnValValAlaAsnArgAlaAspAlaLe 220
Db 4528 GGCCAAAGCACCGGATCTTCGCGGTGCCCGCAGGTGCTGGTGAATCGCGCGCATGCCCT 4587
QY 220 uThrArgProLeuIleAlaAspThrValProCysTyrSerPheGlyLeuAsnLysProAs 240
Db 4588 GACCCGAGCGCTGATCCGATACCGTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4647
QY 240 pPheLysAlaPheGlyLeuIleGluAspGlyGlnLysTrpLeuAlaPheGlnPheAs 260
Db 4648 CTTCAAGGCTTTCGGGCTGATCGAGGAAGACGGCCAGAGTGGCTGGCGTTCACGTTCCG 4707
QY 260 p-LysLeuLeuProValGlyGluLeuLysIleArgGlyAlaHisAsnTyrSerAsnAlaL 280
Db 4708 -CAAGCTGCTGGGTTGGGAACTCAAGATCGTGGCGGCCACACACTATTCCACGCGC 4766
QY 280 euAlaAlaLeuAlaLeuGlyHisAlaValGlyLeuProPheAspAlaMetLeuGlyAlaL 300
Db 4767 TCGCGCGCTGGCGCTGGCGCATCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGC 4826
QY 300 euLysAlaPheSerGlyLeuAlaHisArgCysGlnTrpValArgGluArgGlnGlyValS 320
Db 4827 TGAAGCGGTTTCCGGCGCTGGCTCATCGTGGCAGTGGGTAGCGAGCGGAGGGCGGTGA 4886
QY 320 eTyrTyrAspSerLysAlaThrAsnValGlyAlaAlaLeuAlaAlaIleGluGlyL 340
Db 4887 GCTACTACGACGATTCAGGCCACCAAGCTGGCGCGCGCTGGCGCGGATCGAGGGCG 4946
QY 340 euGlyAlaAspIleAspGlyLysLeuValLeuLeuAlaGlyGlyAspGlyLysGlyAlaA 360
Db 4947 TGGGTGCCGACATCGACGCAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 5006
QY 360 spPheHisAspLeuArgGluProValAlaArgPheCysArgAlaValValLeuLeuGlyA 380
Db 5007 ATTTCCATGACCTGCGCGAGCGCGCTTCTTCTGCGGGCGGTGGTACTGCTTGGCC 5066

```

```

QY 380 rGspAlaGlyLeuIleAlaGlnAlaLeuGlyAsnAlaValProLeuValArgValAlaL 400
Db 5067 GTACGCGCGGCTGATTCGCCAGGCACTGGCAACGGGTACCGTGTGCGCGTCCAA 5126
QY 400 hrLeuAspGluAlaValArgGlnAlaAlaGluLeuAlaArgGluGlyAspAlaValLeuL 420
Db 5127 CGTGGACGAAGCAGTCCGCGAGCCCGAGTGGCCCGGAGGCGATGCGGTGTGT 5186
QY 420 euSerProAlaCysAlaSerLeuAspMetPheLysAsnPheGluGluArgGlyArgLeuP 440
Db 5187 TGTGCGCGGCTCGCGAGCCTTGACATGTTCAAGAACTTCGAAGAAACGCGGACGCTGT 5246
QY 440 healLysAlaValGluGluLeuAla 448
Db 5247 TCGCAAAAGCGTAGAGGAGTAGCG 5272

RESULT 3
AE016779
LOCUS
DEFINITION
Pseudomonas putida KT2440 section 6 of 21 of the complete genome.
ACCESSION
AE016779 AE015451
VERSION
AE016779.1 GI:26557023
KEYWORDS
Pseudomonas putida KT2440
SOURCE
Pseudomonas putida KT2440
ORGANISM
Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
Pseudomonadaceae; Pseudomonas.
REFERENCE
1 (bases 1 to 301995)
Nelson, K., Paulsen, I., Weinel, C., Dodson, R., Hilbert, H., Fouts, D.,
Gill, S., Pop, M., Martins Dos Santos, V., Holmes, M., Brinkac, L.,
Beanan, M., DeBoy, R., Daugherty, S., Kolonay, J., Madupu, R.,
Nelson, W., White, O., Peterson, J., Khouri, H., Hance, I., Lee, P.,
Holtzapple, E., Scanlan, D., Tran, K., Moazzzez, A., Utterback, T.,
Rizzo, M., Lee, K., Kosack, D., Moestl, D., Kiewitz, C., Eisen, J., Timmis, K.,
Hohseil, J., Straetz, M., Helm, S., Kiewitz, C.,
Duesterhoft, A., Tummeler, B. and Fraser, C.
Complete genome sequence and comparative analysis of the
metabolically versatile Pseudomonas putida KT2440
Environ. Microbiol. 4 (12), 799-808 (2002)
2 (bases 1 to 301995)
Nelson, K., Paulsen, I., Weinel, C., Dodson, R., Hilbert, H., Fouts, D.,
Gill, S., Pop, M., Martins Dos Santos, V., Holmes, M., Brinkac, L.,
Beanan, M., DeBoy, R., Daugherty, S., Kolonay, J., Madupu, R.,
Nelson, W., White, O., Peterson, J., Khouri, H., Hance, I., Lee, P.,
Holtzapple, E., Scanlan, D., Tran, K., Moazzzez, A., Utterback, T.,
Rizzo, M., Lee, K., Kosack, D., Moestl, D., Kiewitz, C., Eisen, J., Timmis, K.,
Hohseil, J., Straetz, M., Helm, S., Kiewitz, C.,
Duesterhoft, A., Tummeler, B. and Fraser, C.
Direct Submission
Submitted (05-NOV-2002) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA
FEATURES
source
1. 301995
/organism="Pseudomonas putida KT2440"
/mol_type="genomic DNA"
/strain="KT2440"
/db_xref="taxon:160488"
89. .712
/gene="sspA"
/locus_tag="PP1320"
89. .712
/gene="sspA"
/locus_tag="PP1320"
/note="similar to GP:9864180; identified by sequence
similarity; putative"
/transl_table=11
/product="stringent starvation protein A"
/protein_id="AA066944.1"
/db_xref="GI:24982777"
KLAENPYGSVPTLVDRDLALYESTVVMYLEERYPHPLMPYVPVARGNSRLMHRI

```

```

ORDMCALADTVLDPRSSSEARTKARKALRESLTGVSPLFSBFACFMDSQSLVDCCLL
PITMLPVLGTELPQAKPLLDYMERQAFREPFQASLSVERMKL"
693. .1154
/ gene="sspb"
/ locus_tag="ppl321"
693. .1154
/ gene="sspb"
/ locus_tag="ppl321"
/ note="similar to GB:S50157, GB:X60221, SP:P24539, and
PID:509291; identified by sequence similarity; putative"
/ codon_start=1
/ transl_table=11
/ product="stringent starvation protein B"
/ protein_id="AAN66945.1"
/ db_xref="GI:24982778"
/ translation="MKCASFEKLMSRRPYLVRLALYEMIVDNDCTPHMLYNAEPKV
QVPDGFASDGOIVLINISPSAVRSLSHMDNDVSEFEFGSVAHSLFVPPVGAIIGIYARE
NGQGVFLEFPSSMDGELEDDVQDDGPPDGGQPRPTGRPSLAVK"
complement(1216. .1794)
/ locus_tag="ppl322"
complement(1216. .1794)
/ locus_tag="ppl322"
/ codon_start=1
/ transl_table=11
/ product="lipoprotein, putative"
/ protein_id="AAN66946.1"
/ db_xref="GI:24982779"
/ translation="MTPMRLGLMALTCLSVTGSSVLTSTNSPIEDDRTGTIGSK
IDDSLIETKAANTSKANPNLDKSHIVVSYNGIVLLAGOTPRADLSAEQASQV
QRVKVHNELOVMQPSLSILARNNDNLWTKTKTQMLTDSAVPSSRIKVTENGIVILL
GLVTRQANSATAVQGVGVQKIVKLFYID"
complement(1791. .2378)
/ gene="gmhA"
/ locus_tag="ppl323"
complement(1791. .2378)
/ gene="gmhA"
/ locus_tag="ppl323"
/ note="identified by match to PFAM protein family HMM
PF01380"
/ codon_start=1
/ transl_table=11
/ product="phosphoheptose isomerase"
/ protein_id="AAN66947.1"
/ db_xref="GI:24982780"
/ translation="MQSRIRLFEASITKQOAMDILAPHIEQASLVNMVALLNEGKM
LACNGSGAGAHFSELLRFRERESLPAIALTTDSITSTANDSYNEVFSKO
IRALGQGVLLATISGNSANVIOAIQAHDREMIYVALTGRDGGGNASLLLPEDVE
IRVPTVTRIQEVHLLAIHCLDLIDSQLFSEE"
complement(2471. .2845)
/ locus_tag="ppl324"
complement(2471. .2845)
/ locus_tag="ppl324"
/ note="identified by match to TIGR protein family HMM
TIGR00252"
/ codon_start=1
/ transl_table=11
/ product="conserved hypothetical protein TIGR00252"
/ protein_id="AAN66948.1"
/ db_xref="GI:24982782"
/ translation="MMAASPTRAGQAAETQALEYLOGQLLLARNWRCKGGELDLV
MLDADTVVFEVRYRLHAGFGGALDSIGRKQRLVLAATLFLQEAHWGNYPCRFDV
VALQGSHHAGRPLOWLKNAPFC"
complement(2842. .4659)
/ locus_tag="ppl325"
complement(2842. .4659)
/ locus_tag="ppl325"
/ codon_start=1
/ transl_table=11
/ product="lipoprotein, putative"
/ protein_id="AAN66949.1"
/ db_xref="GI:24982783"
/ translation="MIACLRLLTALCIAALLAACAASSPSSSGLGELPPTPDASIQELLE
KAASSKAEDALLURLSAADLAKYKQKQFPFRAARILEQVPLDTLPAQGVFASTLADEL

```

```

AMSRNQKAALTALAHPSLQRLAEPLADLOARTYSVHAAALPADQALAAAOQRILLA
PLLGQAASTNDNDAIWLAVSLPAEQLOQPTNDQTLGAGTSLAFVKSAGTLEQQQAA
TEAVYKHDPHAPAAQPLATLTKLKELASOPTLKTALLPOBEGPLAGVARLUDQFMA
AHFOAQSGQAPAVQVPPSSRIGSLDDFYBQAOAGVOLVIGPLEKPLVKOLAAPQ
LPITTLALNDADGAKAPPQFQFLAAEDAREVARRARADCMWRANVALVQSGEWD
RVLAARQDWEWGGLTAAERIAQFVALAQIADLFQLOSREGRAKSLQSVTGSIA
AQPSRRQDIDFTLASTPQQAQIIPKPTLNQFYAGDVPIYATSNLYSASGDINQYNDMN
GTRFCETPWLDDTSNLSRQQVQWQPAAGSLGRLYAMGVDAVSLAPRLGQLKALPDN
RVGLSGSLSINANQVERQLPWAEEFGQVKRLPDIVR"
4837. 5721
/ locus_tag="ppl326"
4837. 5721
/ locus_tag="ppl326"
/ note="identified by match to PFAM protein family HMM
PF00590"
/ codon_start=1
/ transl_table=11
/ product="tetracycline methylase family protein"
/ protein_id="AAN66950.1"
/ db_xref="GI:24982784"
/ translation="MCVTDVAGASKSTVGTLVVVATPTGNLDDMSARALKVLADVAL
IAEDTRHSIRLQHFQIDTPLAACHEHNERDEGGRFLSKLLAGENVALVSDAGTPLI
SDPGVHLVQARAAGSVVPPGACALIAALSAAGLPSDRETFEGFLPAKQARRARL
EQVREEPTLIFYRPHRLKLCEDMEAVFGGDRPALLARELTKTFFELKGLPLAELR
AFVAGDSNQQRGECVVLVGGWSAPEGEQAISAEQORVLDLLLAELPLKRAALAAEIT
GVRKNLLYQLALEKQAE"
complement(5744. .6164)
/ locus_tag="ppl327"
/ note="this region contains a gene with one or more
premature stops or frameshifts, and is not the result of a
sequencing artifact; conserved hypothetical protein,
degenerate; similar to GB:X16944, and SP:P15488;
identified by sequence similarity; putative"
6455. 6928
/ locus_tag="ppl328"
6455. 6928
/ locus_tag="ppl328"
/ note="identified by match to PFAM protein family HMM
PF02381"
/ codon_start=1
/ transl_table=11
/ product="conserved hypothetical protein TIGR00242"
/ protein_id="AAN66951.1"
/ db_xref="GI:24982786"
/ translation="MWGITAVRGNANVSLDAKGLAMPSPRYRDELDRCNGQLIVTI
DAVDFCLCVPLDWEQTEAKLRALPSLEENRRLQLLIGNAVLELDGSGRFLVPP
RLREYAKDKKAMLVQLNKFQMDWEDANNAVSAADLAAIQPGAMPDELRLIL"
6925. 7872
/ locus_tag="ppl329"
6925. 7872
/ locus_tag="ppl329"
/ note="similar to GP:9950652; identified by sequence
similarity; putative"
/ codon_start=1
/ transl_table=11
/ product="conserved hypothetical protein TIGR00006"
/ protein_id="AAN66952.1"
/ db_xref="GI:24982787"
/ translation="MTIDSGFNHITVLDEAVEALRADGCYLDGTFGRGGHSRLIL
SKLPGQRLIGFDKDPQAIATQALAAEDGRFVIVQSFAGELGAEGAAGLGHKVSGL
LIDLGVSFQDDDPERGSFLNDGFLDMRNPDCQVSAEETATAPVEIARVFEYEG
EERFAGRNARAVERRERKQPTRTADLAELVKVANPEWKGKNKPNATRAPQGLRIHVN
ELGDLEAGLEAALDALEYGGRLLAVISFHLSDRIYVKLFMRILVKGADNLPRNLVQVH
KVFEPPKIKLIGKAQFASEALKANPRSRSAVMRVAEKL"
7869. 8162
/ gene="ftsL"
/ locus_tag="ppl330"
7869. 8162
/ gene="ftsL"
/ locus_tag="ppl330"
/ note="similar to GB:M11186, GB:M62611, SP:P01178,
PID:L89411, PID:386991, and PID:825699; identified by
sequence similarity; putative"

```

## Alignment Scores:

Pred. No.: 5,36e-24 Length: 301995  
Score: 35.00 Matches: 35  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 7.81% Indels: 0  
DB: 1 Gaps: 0

US-09-701-229-2 (1-448) x AE016779 (1-301995)

QY 136 LysargValAlaValGlyAsnLeuGlyThrProAlaLeuAspLeuLeuAlaAspasp 155  
|||||  
DB 14248 AAACGCGTGGCGTGGCGGTAACTCGTACCGCGACCTGCCTGCGCGATGAC 14307  
|||||  
QY 156 IleGluLeuTyrrValLeuGluLeuSerSerPheGlnLeuGluThr 170  
|||||  
DB 14308 ATCGAGCTGTAGCTGCTGAGCTTTCACGCTCCAGCTGGAAACC 14352  
|||||

## RESULT 4

AE016871/c 311600 bp DNA linear BCT 05-MAR-2003  
LOCUS Pseudomonas syringae pv. tomato str. DC3000 section 16 of 21 of the  
DEFINITION complete genome.

ACCESSION AE016871 AE016853

VERSION AE016871.1 GI:28854552

## KEYWORDS

SOURCE Pseudomonas syringae pv. tomato str. DC3000

## ORGANISM

Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;  
Pseudomonadaceae; Pseudomonas.

## REFERENCE

AUTHORS Buell, R., Joardar, V., Khouri, H., Fedorova, N., Tran, B., Russell, D.,  
Berry, K., Utterback, T., Van Aken, S., Feldblyum, T., Gwinn, M.,  
Dodson, R., DeBoy, R., Durkin, A., Kolonay, J., Madupu, R.,  
Daugherty, S., Brinkac, L., Beanan, M., Haft, D., Selengut, J.,  
Nelson, W., Davidson, T., White, O., Fraser, C. and Collmer, A.  
Complete Sequence of Pseudomonas syringae

## TITLE

Unpublished

## JOURNAL

2 (bases 1 to 311600)

## REFERENCE

AUTHORS Buell, R., Joardar, V., Khouri, H., Fedorova, N., Tran, B., Russell, D.,  
Berry, K., Utterback, T., Van Aken, S., Feldblyum, T., Gwinn, M.,  
Dodson, R., DeBoy, R., Durkin, A., Kolonay, J., Madupu, R.,  
Daugherty, S., Brinkac, L., Beanan, M., Haft, D., Selengut, J.,  
Nelson, W., Davidson, T., White, O., Fraser, C. and Collmer, A.  
Direct Submission

## TITLE

Submitted (03-MAR-2003) The Institute for Genomic Research, 9712

## JOURNAL

Medical Center Dr, Rockville, MD 20850, USA

## FEATURES

## source

1. .311600  
/organism="Pseudomonas syringae pv. tomato str. DC3000"  
/mol\_type="genomic DNA"  
/strain="DC3000"  
/db\_xref="taxon:223283"  
/note="pathovar: tomato"  
131..601

## gene

/gene="bfr"

## CDS

/locus\_tag="PSPT04160"

131..601

/gene="bfr"

/locus\_tag="PSPT04160"  
/note="similar to SP:P22759, and SP:P22759; identified by  
sequence similarity; putative"

## CDS

/codon\_start=1

/transl\_table=11

/product="bacterioferritin"

/protein\_id="AAO57616.1"

/db\_xref="GI:28854553"

/translation="MKGDITVIQHLNKLIGLNLVAINOYFLHARMYEDWGLKLGHGE  
YKESIHAMKADLKIKRTFLGLNPLQDLGKILIGHTKEMLECDLKEOKGLADLK  
AAIAHFTVGGFSGRELLEDILEPKEEQIDWLEQLGLIVKVGNIYLOSQMGD"

## gene

complement(689..1015)

/locus\_tag="PSPT04161"

## CDS

complement(689..1015)  
/locus\_tag="PSPT04161"  
/note="identified by match to TIGR protein family HMM  
TIGR00365"  
/codon\_start=1  
/transl\_table=11  
/product="glutaredoxin-related protein"  
/protein\_id="AAO57617.1"  
/db\_xref="GI:28854554"  
/translation="MDITETIKEQIASNTILLYMKGAPNAPQCQPSAKASQALMACGE  
KFAYVDILQNFPEIRANLPKYNWTFPQWVAGELVGGSDIITENMADGSLQTLVKEA  
SAAKAE"  
complement(1104..3302)  
/locus\_tag="PSPT04162"  
complement(1104..3302)  
/locus\_tag="PSPT04162"  
/note="similar to GP:5441785; identified by sequence  
similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="oxidoreductase, molybdopterin-binding"  
/protein\_id="AAO57618.1"  
/db\_xref="GI:28854555"

## gene

/translation="MARVIARHSHCSAIHRIAMTRPEAMTRTOHLRACHICEAICG  
LAIETVTEPGAAPKITSIKGDLPTFSRGHICPKAVALQDIONDRLRQMLRTGDQ  
WQIEWQAADFVAERLYAIQQHQGNAYVYQGNPSVHNYGLMTHSNYFLGLKTRN  
RFSATSVDPHLHLSFLMYGHGMLLPIDIDHTDFMLILGNPLASNGSIMTVPDVE  
KRLKAIQORGKLVVDPRSETAIAIDQHVFPREGDAALLGLLNTLFEGLTRES  
HLPVDGLHVRHSIAGTSAEAMSPRCGLAAYQIROLARDFAGATLCYGRMGVSTQA  
FGTGLHQAQLINLVNLDNRVGGTLCPEAVDLVSSGGHFNWQVRSGLPEYGG  
ELPVSALAEMLVGGQGVRAVTVAGNPVLSITPNRGDQLALEGLEMLSIDLYIE  
TTRHADLILPSTSALENDHYDTNTLAVRNVTRFNRAIFDKPEGALHDLWETFGVLA  
SFAKAGRELKTPVPAQVIMIDVLRAGRYGAASPHNLSLETGLASHPHGVMDGALKPNL  
DRLKTAQRIOAAPEVTMIDVLRAGRYGAASPHNLSLETGLASHPHGVMDGALKPNL  
HRLVGRPHQLLHPDDLHRGLSDGQQVRSRVGVIEVYELGSLDMPGVSLPH  
GWGHSRCVMEIARNPQVSAANDLTDRLQDLDELGNALNGVPVQVASC"

## gene

complement(3357..3650)  
/locus\_tag="PSPT04163"  
complement(3357..3650)  
/locus\_tag="PSPT04163"  
/note="identified by Glimmer2; putative"  
/codon\_start=1  
/transl\_table=11  
/product="conserved hypothetical protein"  
/protein\_id="AAO57619.1"  
/db\_xref="GI:28854556"  
/translation="MDSKEHQLTELLGLTARTLTLTASMTSMPELLRSDEVTRNA  
GRMDRMATISSGLDEHWRLLGLDTGTHMAQEYVETVTEIQLQRKIYTPGG"

## gene

4100..5020  
/gene="argF"  
/locus\_tag="PSPT04164"  
4100..5020  
/gene="argF"  
/locus\_tag="PSPT04164"  
/note="identified by match to PFAM protein family HMM  
PF02729"

## CDS

/codon\_start=1  
/transl\_table=11  
/product="ornithine carbamoyltransferase"  
/protein\_id="AAO57620.1"  
/db\_xref="GI:28854557"  
/translation="MNAHFSLMMDYTPDELGLIRRGVELKDLNRNGLVFEPLKNRY  
LGMIFKSTRRLSFEAGMQLGQQAIFLSHRDTQLGRGEPIDASAKVISMLDAVM  
IRTVADHNLTEFAANSRVPVINGLSDLLCOLLADMTQFLFHRGSIKGKTAVTGDG  
NMCNSYIEAAIOEPQFOLRVACPGAYEPNPELALAGERTVVRDPKAAVAGALVST  
DYTSMGQEEETARKKALFAPQVTRALLDLADKVDLFWHCLPAHRGEEISVDLLDDA  
RSVAWDAENRHLHAQKALLEFLVAPCFQPA"

## gene

5026..6087  
/locus\_tag="PSPT04165"

## CDS

complement(689..1015)  
/locus\_tag="PSPT04165"  
/note="similar to GB:248042, SP:Q14444, and PID:662994;  
identified by sequence similarity; putative"

```

/codon_start=1
/transl_table=11
/product="ABC transporter, ATP-binding protein"
/protein_id="AA057621.1"
/db_xref="GI:2885458"
/translation="MLNLRNLACGYQNRVVDLNLHLNAGDITGCLLGGSCGKTTT
LRAITAGETIHAGETIAGDVISRPQWTLAPKRRIGWVFODYALFPHLSVADNIAG
IQPHLDVIGLELVNLCALGKRVPHLSGGQQORVALARALAPEPQLLLIDEPF
SNLDELRRRLSHVRDILKARGTSAILVYHDQEEAFVSDHGVGFRRLEQMDTPY
NLYHEPQTFFVASTIGQYFIRGQMIPEPSVHELVLRGNRPYVGRGSAVDVLLRP
DDIVYAPSDLKARVIGRTFKGASTLYRLQLPTGSLQLEAIFPSSHADHRQGEVGI
ADHLVFAAGSVAAYAG"
/complement(6036..6566)
/locus_tag="PSPT04166"
/complement(6036..6566)
/locus_tag="PSPT04166"
/notes="similar to SP:P21484; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="ybaK/ebcC protein"
/protein_id="AA057622.1"
/db_xref="GI:28854559"
/translation="MTPALDLLKKNRAEHRHSYHDPKAAASYGLEAAEKGLDAA
FKTLASTERKGLLVAVPVVGTLDLALANAAGAKKTENADPAARQSRGYLLGGIS
PLGQKKRLRTFIDETALESIVSAGRGLEVELTAVALAETHQARFAPIGRE"
6918..7775
/genes="glpP"
/locus_tag="PSPT04167"
6918..7775
/genes="glpP"
/locus_tag="PSPT04167"
/notes="similar to SP:P11244; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="glycerol uptake facilitator protein"
/protein_id="AA057623.1"
/db_xref="GI:28854560"
/translation="MTIALKQPTLCQCAEFLGTALMIFFGTGCVAALKVAGATFGL
WEICLWMAVSMGIVLSAGISGAHLNPAVSIASLFAFGRKLPFYISAQIAGAF
GAGLVLYLISDFEHAHRIIRSEQLASVFTYPPNPAISVQAFLEVFWIT
ILMGVIMALDDNNFGRALAPLLGLLVAVIGSAMGPLTGFAMPNAPDFGPKMTF
FAGWGIEATFGGRDIPYFLVPIFAPILGACLAGAARYALRIARLHPSTVTVNEKEAPV
VRGKVOASS"
7805..9310
/genes="glpK"
/locus_tag="PSPT04168"
7805..9310
/genes="glpK"
/locus_tag="PSPT04168"
/notes="identified by match to PFAM protein family HM
PF02782"
/codon_start=1
/transl_table=11
/product="glycerol kinase"
/protein_id="AA057624.1"
/db_xref="GI:28854561"
/translation="MTDTQNKYIIIALDQGTSSRAIIFDRDANVYVSTAQSEFVQHP
QAGVHEDPMEIFATQCTMKAQADLHQDIAAIGITNQRETTVIERDTGRPIY
NAIVQCRSTEQQLKRDGLEEVTKDTGLVLDLPYFSGSKVKILDHVGSRRERAR
KGLMFGITDWLWKFTGGKVHVVDYNSRTMLFNHTLEWDQRMDLVDLPREIL
PEVKSSEVYGHKSQIPITAGIAGQQAALFGQMCVEFGQAKNTGTGCLLMNTGKK
AVKSHAGLTTICGPRGEVATAGLGVNGSTVQWLRDELKLNLDLDTYEFVAGY
KDSNGYILPAPTGLGAPYDPYARGALFGLTRGVKVDHIIIRAALESAYQTRDVLDA
MOQDSGERLKSURVDGAVANNFLMQFDILGTHVERPQMRRTALGAFLAGLAIG
FWSSLDLRNKAVIERVEFACEAHEKLYAGWKAVARTDWEPHENE"
9421..10176
/genes="glpP"
/locus_tag="PSPT04169"
9421..10176
/genes="glpP"
/locus_tag="PSPT04169"

```

/note="similar to SP:Q51391; identified by sequence similarity; putative"

Alignment Scores:  
Pred. No.: 2,33e-08 Length: 311600  
Score: 20.00 Matches: 20  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 4.46% Indels: 0  
DB: 1 Gaps: 0

US-09-701-229-2 (1-448) x AE016871 (1-311600)

Qy 278 AsnAlaLeuAlaLeuAlaLeuGlyHisAlaValGlyLeuProPheAspAlaMetIeu 297  
|||||  
Db 291021 AACGACATGGCAGCCCTGGCGCTGGCGCATGCCCTGGCGCTTCCGACCCATGCTT 290962  
|||||

#### RESULT 5

AX067460/c AX067460 96109 bp DNA linear PAT 24-JAN-2001  
LOCUS AX067460  
DEFINITION Sequence 35 from Patent WO0078968.  
ACCESSION AX067460  
VERSION AX067460.1 GI:12545080  
KEYWORDS Moraxella catarrhalis  
SOURCE Moraxella catarrhalis  
ORGANISM Moraxella catarrhalis  
Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;  
Moraxellaceae; Moraxella.

#### REFERENCE

1 Lagace, R.E., Patterson, C. and Berg, K.L.  
AUTHORS Nucleotide sequences of moraxella catarrhalis genome  
TITLE Patent: WO 0078968-A 35 28-DEC-2000;  
JOURNAL Incyte Genomics, Inc. (US)

#### FEATURES

source  
1..96109  
/organism="Moraxella catarrhalis"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:480"

BASE COUNT 28783 a 18910 c 20341 g 28075 t  
ORIGIN

#### Alignment Scores:

Pred. No.: 8.96e-08 Length: 96109  
Score: 19.00 Matches: 19  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 4.24% Indels: 0  
DB: 6 Gaps: 0

US-09-701-229-2 (1-448) x AX067460 (1-96109)

Qy 113 AlaIleThrGlySerAsnAlaLysSerThrValThrThrLeuValGlyGluMetAla 131  
|||||

Db 12400 GCCATCACAGGCTCAATGCCAAGACACTGTTACACATTAGTAGCGGAGATGGCA 12344  
|||||

#### RESULT 6

AX189058

LOCUS AX189058 1317 bp DNA linear PAT 08-AUG-2001  
DEFINITION Sequence 259 from Patent WO0148209.  
ACCESSION AX189058  
VERSION AX189058.1 GI:15142764  
KEYWORDS Escherichia coli  
SOURCE Escherichia coli

Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
Enterobacteriaceae; Escherichia.

#### REFERENCE

1 Forsyth, R.A., Ohlsen, K.L. and Zyskind, J.W.  
AUTHORS Genes identified as required for proliferation of E. coli  
TITLE Patent: WO 0148209-A 259 05-JUL-2001;  
JOURNAL Elitra Pharmaceuticals, Inc. (US)  
FEATURES Location/Qualifiers  
source 1..1317



```

    /note="pot. ribosome binding site"
    111 1193

```

/note="ORF-Y (AA 1-360)"		PUBMED		12384590	
/codon_start=1		REFERENCE		2 (bases 1 to 11441)	
/transl_table=11		AUTHORS		Jin,Q., Shen,Y., Wang,J.H., Liu,H., Yang,J., Yang,F., Zhang,X.B., Zhang,J.Y., Yang,G.W., Wu,H.T., Dong,J., Sun,L.L., Xue,Y., Zhao,A.L., Gao,Y.S., Zhu,J.P., Chen,S.X., Yao,Z.J., Wang,Y., Lu,W.C., Qiang,B.Q., Wen,Y.M. and Hou,Y.D.	
/db_xref="GI:42059"		TITLE		Direct Submission	
/transl_table=11		JOURNAL		Submitted (08-MAY-2001) Microbial Genome Center of Chinese Ministry of Public Health, 100 YingXin Jie, XuanWu Qu, Beijing 100052, P.R. China	
/translation="MLVLAHLVKKYSGNVPSLTFRAIVSLTLTALFSLWNGPRM IAHQLKSFQGVNRDGPESHFSKRGPTTGGIMLTALIVSLLWYAPSNPYVCVL VLVGYGVIGFVDYDVRKDKTKGLIARKWKYFWMVSIALGVAFALYIAGKDPATOL VVPEFKDMPQLGLFYLLAYFVIVGTGNVNLTDGLDGLAIMPTEVFVAGGFALVAVA LGTIAVLLEOFEFLVINGGVVETLSVLVQVSGFKLRGQIRFMAPIHIIHHYELKGHP EPRVIVFNLISMLVLGLATLKV"		FEATURES		Location/Qualifiers	
1018..1023		promoter		1..11441	
/note="pot. -35 region"				/organism="Shigella flexneri 2a str. 301"	
1043..1048		promoter		/mol_type="genomic DNA"	
/note="pot. -10 region"				/strain="301"	
1181..1185		RBS		/serotype="2a"	
/note="pot. ribosome binding site"				/db_xref="taxon:198214"	
1196..2512		CDS		292..786	
/note="murD gene product (AA 1-438)"				/gene="yabB"	
/codon_start=1				/locus_tag="SF0078"	
/transl_table=11				292..786	
/protein_id="CAA35933.1"				/gene="yabB"	
/db_xref="GI:42060"				/locus_tag="SF0078"	
/translation="MADYQKNVVIIGLGTGLSCVDFFLARGVTPRVMDTRMTPPGL DKLPEAVERHTSLNDEWLMADLIIVASPGIALHPSLSAAADAGIEIVGDIELFCRE AQAPIVAITSGNSKSTVTLVGMKAAGVNVGGNIGLPALMLDDECELVLELS SFOLETTSSIQAAVAATILNVTEHDMDRYPFGLOQYRAAKLRIYENAKVCVNNADALT MPTRGADERCVSFGVNGDYHLNHQOGETWLRVKGKLVNKKMKLSGQNVNNAALAA LALADAGLPRASLKAITFTTGLPHRFVLEHNGVRVNMKATLRNVGSTEALNGL HVDTLHLGGGKGADFSPLARYLNDVNRVYCFGRDGAQALALRPEVAEQETME QAMRLAPRVQPDGMVLLSPACASLPQFKNFQGRNEFARLAKELG"				/note="Residues 13 to 164 of 164 are 99 pct identical to residues 1 to 152 of a 152 aa protein from Escherichia coli O157:H7 EDL933 ref: NP_285777.1"	
530 a 631 c 777 g 670 t				/codon_start=1	
BASE COUNT				/transl_table=11	
ORIGIN				/product="orf, conserved hypothetical protein"	
Alignment Scores:				/protein_id="AA41743.1"	
Pred. No.:		0.000579		/db_xref="GI:24050284"	
Score:		14.00		/translation="MMGVKGIERGEAGMFGATLVNLDKGRLSVPTRYRQELLENAAG QMCTIDIHPCLLYLPPEWIEIQKLSRLSNPVRVQRLLLGHASECOMDGAG RLVIAPVLRHAGLTKEVLMVGQFNKELWDETTHQQVKEDIDAEQLATGDLSERLQ DLSL"	
Percent Similarity:		100.00%		788..1729	
Best Local Similarity:		100.00%		/gene="yabC"	
Query Match:		3.12%		/locus_tag="SF0079"	
DB:		1.		/note="Residues 1 to 313 of 313 are 99 pct identical to residues 34 to 346 of a 346 aa protein from Escherichia coli pir: QOECF"	
US-09-701-229-2 (1-448) x ECMUROY (1-2608)				/codon_start=1	
Qy		157 GluLeuTyrValLeuGluLeuSerSerPheGlnLeuGluThr 170		/transl_table=11	
Db		1652 GAACGTGACGTGCTGGAACGTGTCGAGCTCCACCTCGAACC 1693		/product="yabC protein"	
RESULT 9				/protein_id="AA41744.1"	
AE015046				/db_xref="GI:24050285"	
LOCUS				/translation="MMENYKHTTVLLDEAVNGLINIRPDGIYIDGTFGRGHSRLILSQ LGEGRLLAIDRDPQAIYAKVTIDDPFRIIHGPFSGALGEYVAERDLKIGIDILDL GVSSPOLDDAERGFSFMRDGPDMRMDPTRGSAAEWLQTAEEADIAWLKTYGEEHF AKRIARAIVERNREQMTRTKELAEVVAATPVKDKFKHPATRTFOAIVTNVNSELEE IQALKSSNLVLAAPGGRLSIISFHSLEDRIKREMSRSGQVPAGLPWTERQLAKL GRQLRALKGLMPGEEVAENPARSSVLRIARTNA"	
DEFINITION				1726..2091	
VERSION				/gene="ftsL"	
KEYWORDS				/locus_tag="SF0080"	
SOURCE				1726..2091	
Shigella flexneri 2a str. 301				/gene="ftsL"	
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Shigella.				/locus_tag="SF0080"	
1 (bases 1 to 11441)				/note="Residues 1 to 121 of 121 are 100 pct identical to residues 1 to 121 of a 121 aa protein from Escherichia coli O157:H7 EDL933 ref: NP_285779.1"	
Jin,Q., Yuan,Z.H., Xu,J.G., Wang,Y., Shen,Y., Lu,W.C., Wang,J.H., Liu,H., Yang,J., Qu,D., Zhang,X.B., Zhang,J.Y., Yang,G.W., Wu,H.T., Dong,J., Sun,L.L., Xue,Y., Zhao,A.L., Gao,Y.S., Zhu,J.P., Kan,B., Chen,S.X., Yao,Z.J., He,B.K., Chen,R.S., Ma,D.L., Qiang,B.Q., Wen,Y.M., Hou,Y.D. and Yu,J.				/codon_start=1	
Genome sequence of Shigella flexneri 2a: insights into pathogenicity through comparison with genomes of Escherichia coli K12 and O157				/product="cell division protein; ingrowth of wall at septum"	
Nucleic Acids Res. 30 (20), 4432-4441 (2002)				/protein_id="AA41745.1"	
TITLE				/db_xref="GI:24050286"	
JOURNAL				/translation="MISRVTEALSVKVSGMSHERHALPGVIGDDLDRFGKLPCLCFI CIIITAVTVVTTAHTRLTLTAQREQVLRLERDALDIERNLILENALGDHSRVERIAT EXLQMHVDPSQENIVVQK"	



Science 269 (5223), 496-512 (1995)  
95350630  
PUBMED  
7542800

2 (bases 1 to 11498)  
Tatusov,R.L., Mushegian,A.R., Bork,P., Brown,N.P., Hayes,W.S.,  
Borodovsky,M., Rudd,K.E. and Koonin,E.V.  
Metabolism and evolution of Haemophilus influenzae deduced from a  
whole-genome comparison with Escherichia coli  
Curr. Biol. 6 (3), 279-291 (1996)

JOURNAL MEDLINE  
PUBMED  
96398784

3 (bases 1 to 11498)  
White,O., Clayton,R.A., Kerlavage,A.R. and Fleischmann,R.D.  
Direct Submission  
Submitted (25-JUL-1995) The Institute for Genomic Research, 9712  
Medical Center Dr, Rockville, MD 20850, USA

JOURNAL MEDLINE  
PUBMED  
8805245

4 (bases 1 to 11498)  
White,O., Clayton,R.A., Kerlavage,A.R. and Fleischmann,R.D.  
Direct Submission  
Submitted (27-SEP-1997) The Institute for Genomic Research, 9712  
Medical Center Dr, Rockville, MD 20850, USA  
The H. influenzae sequence has been updated by R. Fleischmann. New  
database matches have been assigned, product names have been  
improved, and a number of frame shifts have been corrected. We  
gratefully acknowledge the work of Tatusov et al. We have  
incorporated their annotation into the /notes fields of the  
corresponding H. influenzae genes

JOURNAL MEDLINE  
PUBMED  
96398784

5 (bases 1 to 11498)  
White,O., Clayton,R.A., Kerlavage,A.R., Fleischmann,R.D.,  
Peterson,J., Hickey,F., Dodson,R. and Gwinn,M.  
Direct Submission  
Submitted (28-MAY-1998) The Institute for Genomic Research, 9712  
Medical Center Dr, Rockville, MD 20850, USA  
The whole genome was shifted by 588 nucleotides for a new start  
On Oct 1, 1996 this sequence version replaced gi:1221887.

JOURNAL MEDLINE  
PUBMED  
96398784

Location/Qualifiers  
1..11498  
/organism="Haemophilus influenzae Rd"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:71421"  
131..586  
/gene="HI1129"

JOURNAL MEDLINE  
PUBMED  
96398784

131..586  
/gene="HI1129"  
/note="similar to SP:P22186 GB:X52063 PID:40849 PID:42319  
GB:U00096 percent identity: 59.21; identified by sequence  
similarity: putative"  
/codon\_start=1  
/transl\_table=11  
/product="conserved hypothetical protein"  
/protein\_id="AAC22784.1"  
/db\_xref="GI:1574684"

JOURNAL MEDLINE  
PUBMED  
96398784

/translation="MFGATAVNLDKSGRAIVTRVRAETLEKNQGMVCTVDIROS  
LLLYPLDEWEKIEOKLLALSNDFTQRRLORVMLGHTATEMDAQRILLSPLRQHA  
KLEGLMLVGQLNFESWDEWHQTAEIDIEIGSSDFAADALNDFSL"  
617..1382  
/gene="HI1130"  
617..1582  
/gene="HI1130"

JOURNAL MEDLINE  
PUBMED  
96398784

/note="similar to GB:D10483 SP:P18595 GB:X52063 PID:216497  
PID:40850 percent identity: 62.09; identified by sequence  
similarity: putative"  
/codon\_start=1  
/transl\_table=11  
/product="conserved hypothetical protein"  
/protein\_id="AAC22785.1"  
/db\_xref="GI:1574685"

JOURNAL MEDLINE  
PUBMED  
96398784

/translation="MNSSEFSFSSSEHITVLLHEAVNLGALKENGIIYIDGTFRGGHSR  
FILSQLSNGRLIAVDRDPRAIAEAHKIQDLRFQIEHNSFSHIPICDKLNLGVKIDG  
ILLDLGVSSPOLDEARGSFQMGPLDMRDMTQGLSAEEMLKOVSIEDLNVLKTF  
GEERFAKRIATAIVDNKSAVKNGTEFLSTSQAELISQAVPFDDKKHPATRSFQA  
IRIFINSELDESLLSALDMLAPGRUSILIFSLEDNRWMMFKMKOSKGEDIPKG  
LPREDQIQRNKURIKIQAQDAEIGASPRSRAILRVARI"

JOURNAL MEDLINE  
PUBMED  
96398784

1585..1908  
/gene="HI1131"  
1585..1908  
/gene="HI1131"  
/note="similar to SP:P22187 GB:K00137 GB:S49802 GB:S49875  
GB:X52063 percent identity: 36.63; identified by sequence  
similarity: putative"  
/codon\_start=1  
/transl\_table=11  
/product="cell division protein (ftsL)"  
/protein\_id="AAC22786.1"  
/db\_xref="GI:1574686"

JOURNAL MEDLINE  
PUBMED  
96398784

/translation="MSENKPRYPLOQIIVLFSSNKLVLLVLLIGLVSAMGTIWT  
HKTQLISENCMLILQRALENEYRNQVQEATGDSTRVESTAI STLKMKVSSSQRE  
VEIRE"  
1921..3753  
/gene="HI1132"  
1921..3753  
/gene="HI1132"  
/note="similar to GB:D10483 SP:P04286 GB:K00137 GB:S49802  
GB:S49875 percent identity: 53.11; identified by sequence  
similarity: putative"  
/codon\_start=1  
/transl\_table=11  
/product="penicillin-binding protein 3 (ftsI)"  
/protein\_id="AAC22787.1"  
/db\_xref="GI:1574687"

JOURNAL MEDLINE  
PUBMED  
96398784

/translation="MVFNSRSKSGSKTKIRLTAPETVKQNPKPVFEKCFMRGRY  
MLSTVLIIDLGCALVARAAVQSINADTLSEADKRSLRRDEVLSVRGSILDRNQGLL  
SVSPMSAIVADPKTMLUKENSLADKERIALAAEEELGMTENDLVKTERKSKGYLYLA  
RQVELSNANTYIRRLKIIGILETHRRFPVVEEAHVGYTDIDNGTEGIEKSFNS  
LLVGDSRTVRKDRGNIVAHSIDRKKYDAQDVTSIDEKLQSMYRIKKAIVENN  
AESTAVLDVRTGEVLAMATPSYNPNRVGVKSELNRATITDFFEPGSTVKKPVP  
LTALQRCVGRDELIDTSEPLSGKEIVDVAPRAQDTLDEILMNSSNRGVSLRALMP  
PSALMEYQNAGLSKPTDLGLEQVGLNNRWADIERATVAIGITATPPQIA  
RAYATLGSFGYIRPLYSITKYVDPPIGVKRFSEKITKDIVGLEKVAIKNRAMPVEGY  
VGKTTGARKEINGHYNYKVAFTAGTAPISDPYALVLLINDPKAGEYIGGAVSAPY  
FSNTGWYALRANAIPQDAEAAENTTTKSARKRIVIGEHNKNVN"  
3763..5229  
/gene="HI1133"  
3763..5229  
/gene="HI1133"  
/note="similar to GB:D10483 SP:P22188 GB:X55814 PID:285768  
PID:581032 percent identity: 55.72; identified by sequence  
similarity: putative"  
/codon\_start=1  
/transl\_table=11  
/product="UDP-N-acetylmuramyl-tripeptide synthetase  
(murF)"  
/protein\_id="AAC22788.1"  
/db\_xref="GI:1574688"

JOURNAL MEDLINE  
PUBMED  
96398784

/translation="MKKLTALFNLPKNDIELHNVLDSRVKAGDLFVALTKGHQVD  
GNQPIDSALHGSAVVSETSELSHELTVAFIGNVVPVKYQIHLAHLSADVDFVTGSV  
SNNTLVGVGTNGKTTISQLQAWAELLGHRAAVNGTIGNGLFOIVAKRNTTGTSV  
EI0SSSAFKHAGADETSESSSHGAQHREALHFKAIFTNLRDHLDYHQSMENY  
AAAKRLFTELDQIKVINADDEIGYQWLTPDAIVASMNADDKVGSHQWMKAINIH  
YHFKGADITPESSWGNGVLSPGIGAFTVSNLLVMTLLSFYGPLENLTAFAKSLKG  
VCGRMEMIOVPNKPTVIDYAHTPDPALEKALLAAREHCQGLWCIFGCGGDRGRKP  
LMAQVAFQAEKLIIVKDNRPRESQSIETDIVAGFKNMKEVGIILPDRAQALQFAIES  
AVENDVTLIAGKHGEHYQIIGSEVHFSDQETALDFLK"  
5243..6616  
/gene="HI1134"  
5243..6616  
/gene="HI1134"  
/note="similar to GB:D10483 SP:P11980 PID:216500 PID:40854  
PID:42048 percent identity: 51.88; identified by sequence  
similarity: putative"  
/codon\_start=1  
/transl\_table=11  
/product="UDP-murNAC-pentapeptide synthetase (murF)"  
/protein\_id="AAC22789.1"  
/db\_xref="GI:1574689"

JOURNAL MEDLINE  
PUBMED  
96398784

/translation="MIKLSVTQALQTLQAKLIGDENVOVEKINTDTRKSVSNSLFFAL"

KGERFDAQYLDQAVSOGALVAVQOENSSISVPQLVVKDTRIALGELAKWLREKINP  
RTVANGSSGTTTKENTASITLOHTAADSEAVLFNTNGFNNDIGVPLTLRLTEKHRF  
AIVELGANOHEINYYTKLVQPNAAITNAPAHLEFGVSLAGVVOAKGYIRGLTKRN  
GVATINAEHLDDWQKEISNHAIOYENGKDYSAKNTHHTSOGSTFTLISPOGRIET  
LPYLGEHVNALATAMVGNLTLDVKGLEQORSQVKGRLFPPIQVTPNLLILDDT  
LYNANKSICRAIDVLKGYDAPRILCVGDMKELGENSLAIHREVGQYINLVNDDLVCSY  
GNESAVTSEAVSGKHFTDKTEMVDPLVPLIENQLQNNKVVVLGKSRSMKMDVIYS  
LKDKIKC

gene

6610.7692

CDS

/gene="H1135"

6610.7692

/gene="H1135"

/note="similar to GB:D10483 SP:PI5876 GB:X51584 PID:216501  
PID:40855 percent identity: 76.67; identified by sequence  
similarity: putative"

/codon\_start=1  
/transl\_table=11  
/product="phospho-N-acetylmuramoyl-pentapeptide-  
transferase E (mray)"  
/protein\_id="AAC22790.1"

/db\_xref="GI:1574690"

/translation="MLVLAELVRYETAFAISYITVRANLALLTALFISLWIGPKV  
IKRLQILKFGQEVNDPESHFAKCTPTMGGMVILFSGVSTLLNANLPIYIWLCL  
FVLEFGAIGFVDDFRKTRNTDGLTGARNVFNWVVALVAILMLVWLGADTDATRL  
VIPEFKDIMPQLGLFYIVLSYFVITGNAVNLTDGLDGLAIMTLVAGAFALIAWA  
TGNVFAEYLHIPIKYSSVVEFTAIAGSGLFGLWFTYPAQVFMGDVGSALGGA  
LGVVAILVROEFLVIMGVFVVEALSIVLQVGSYKLKQKQIFRMADTHHFFELKGWP  
EPVYIIRFWIISLMVLMLGLVTLKL"

gene

7815.9128

CDS

/gene="H1136"

7815.9128

/gene="H1136"

Alignment Scores:

Pred. No.: 0.00218 Length: 11498  
Score: 14.00 Matches: 14  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 3.12% Indels: 0  
DB: 1 Gaps: 0

US-09-701-229-2 (1-448) x U32793 (1-11498)

Qy 157 GluLeuTyrValLeuGluLeuSerPheGlnLeuGluThr 170

Db 8271 GAACCTTATGCTAGAGCTTCTAGTTTTCAGCTTGAGACA 8312

RESULT 11

AX191762

LOCUS

DEFINITION

AX191762

ACCESSION

AX191762.1

VERSION

GI:15209931

KEYWORDS

Haemophilus influenzae

Haemophilus influenzae

Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;

Pasteurellaceae; Haemophilus.

Iversen, P.L.

Antisense antibacterial cell division composition and method

Patent: WO 0149775-A 44 12-JUL-2001;

Avi Biopharma, Inc. (US)

Location/Qualifiers

1..11498

/organism="Haemophilus influenzae"

/mol\_type="genomic DNA"

/db\_xref="taxon:727"

3446 a 1842 c 2509 g 3701 t

BASE COUNT

ORIGIN

Alignment Scores:

Pred. No.: 0.00218 Length: 11498

Score: 14.00 Matches: 14  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 3.12% Indels: 0  
DB: 6 Gaps: 0

US-09-701-229-2 (1-448) x AX191762 (1-11498)

Qy 157 GluLeuTyrValLeuGluLeuSerPheGlnLeuGluThr 170

Db 8271 GAACCTTATGCTAGAGCTTCTAGTTTTCAGCTTGAGACA 8312

RESULT 12

AE005185

LOCUS

DEFINITION

AE005185

VERSION

AE005185.1

KEYWORDS

SOURCE

ORGANISM

Escherichia coli O157:H7 EDL933

Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;

Enterobacteriaceae; Escherichia.

REFERENCE

1 (bases 1 to 12791)

PERNA, N.T., PLUNKETT, G. III, BURLAND, V., MAU, B., GLASNER, J.D.,

ROSE, D.J., MAYHEW, G.F., EVANS, P.S., GREGOR, J., KIRKPATRICK, H.A.,

POSFAL, G., HACKETT, J., KLINK, S., BOUTIN, A., SHAO, Y., MILLER, L.,

GROTBACH, E.J., DAVIS, N.W., LIM, A., DIMALANTA, E., POTAMOUSIS, K.,

APODACA, J., ANANTHARAMAN, T.S., LIN, J., YEN, G., SCHWARTZ, D.C.,

WEICH, R.A. and BLATTNER, F.R.

Genome sequence of enterohaemorrhagic Escherichia coli O157:H7

Nature 409 (6819), 529-533 (2001)

21074935

PUBMED

11206551

2 (bases 1 to 12791)

PERNA, N.T., PLUNKETT, G. III, BURLAND, V., MAU, B., GLASNER, J.D.,

ROSE, D.J., MAYHEW, G.F., EVANS, P.S., GREGOR, J., KIRKPATRICK, H.A.,

POSFAL, G., HACKETT, J., KLINK, S., BOUTIN, A., SHAO, Y., MILLER, L.,

GROTBACH, E.J., DAVIS, N.W., LIM, A., DIMALANTA, E., POTAMOUSIS, K.,

APODACA, J., ANANTHARAMAN, T.S., LIN, J., YEN, G., SCHWARTZ, D.C.,

WEICH, R.A. and BLATTNER, F.R.

Direct Submission

Submitted (22-OCT-2000) Laboratory of Genetics, University of

Wisconsin, 445 Henry Mall, Madison, WI 53706, USA

Location/Qualifiers

1..12791

/organism="Escherichia coli O157:H7 EDL933"

/mol\_type="genomic DNA"

/strain="EDL933"

/serotype="O157:H7"

/db\_xref="taxon:155864"

/note="enterohaemorrhagic"

72..1076

/gene="fruR"

/note="synonym: 20090"

72..1076

/gene="fruR"

/function="regulator; Energy metabolism, carbon:

Glycolysis"

/note="Residues 1 to 334 of 334 are 100.00 pct identical

to residues 1 to 334 of 334 from Escherichia coli K-12

Strain MG1655; B0080"

/codon\_start=1

/transl\_table=11

/product="transcriptional repressor of fru operon and

others"

/protein\_id="AAG54384.1"

/db\_xref="GI:12512783"

/translation="MKLDEIARLAGVSRRTSYINGKAOYRVSDKTVKVMVYRE

HNYPHNAAGIRAGRTRISIGLIVPDLENTSYRIANYLERQAOBGYQLIACSEDO

PDNEMRCIEHLQROVDRIIVSTSLPPEHPFYQWANDPFFIVALDRALDRHFTSVV

GADQDDAEMLAEELRKFPATYLVILGALFELSVSFLREQGFTAWKDDPPEVHFILYAN

```

SYREAAQLEFEKLETHMPQALFTTTFALQGVMDVTLRRDGKLPDSLAIATFGDN
ELLDFQCPVLAAQRHRDVAERVLEIVLASLDEPRKPKGLTRIKRNLRRGVLSRS
1678. 2136
/ gene="yabB"
/ note="synonym: Z0091"
1678. 2136
/ gene="yabB"
/ function="orf; Unknown function"
/ note="Residues 1 to 152 of 152 are 99.34 pct identical to
residues 1 to 152 of 152 from Escherichia coli K-12 Strain
MG1655; B0081"
/ codon_start=1
/ transl_table=11
/ product="orf, hypothetical protein"
/ protein_id="AAG54385.1"
/ db_xref="GI:12512784"
/ translation="MFRGATLVNLSKGRSLVPTRYREQLLENNAQOMVCTIDIHHPG
LLYPLPELIEIKSLRSLSSMPVRRVORLLGHASEQMDGAGRLIATPYLRQHA
GLTKEVNLVGGQFNKFLMDWETTHQOVKEDIDAEQLATGDLSERLQDSL"
2138. 3079
/ gene="yabC"
/ note="synonym: Z0092"
2138. 3079
/ gene="yabC"
/ function="orf; Not classified"
/ note="Residues 1 to 313 of 313 are 100.00 pct identical
to residues 1 to 313 of 313 from Escherichia coli K-12
Strain MG1655; B0082"
/ codon_start=1
/ transl_table=11
/ product="putative apolipoprotein"
/ protein_id="AAG54386.1"
/ db_xref="GI:12512785"
/ translation="WMENYKHTTVLLDEAVNGLNIRPDGIYIDGTFGRGGHSLILSQ
LGEGRDLAIDRDPQAIQAVKIDDPRESIHGFSALGEYVAERDLGKIDGILLDL
GVSSQLDDAERGFSPMRGDPIDMRDPTRGQSAENLQTAEEADIAWVLYKYGGEF
AKRIARIVERNREQPMRTKELAEVAAATQKDKFKHPATRTFOAVRIWYNSELEE
IQALKSNLAPGRLSIISFSLSDRIKVRFNRENSRGQVQVAGLPMTBEQLKLL
GKQRLAKGLMPGEEVAENPRARSSVLRAERTNA"
3076. 3441
/ gene="ftsL"
/ note="synonym: Z0093"
3076. 3441
/ gene="ftsL"
/ function="phenotype; Cell division"
/ note="Residues 1 to 121 of 121 are 100.00 pct identical
to residues 1 to 121 of 121 from Escherichia coli K-12
Strain MG1655; B0083"
/ codon_start=1
/ transl_table=11
/ product="cell division protein; ingrowth of wall at
septum"
/ protein_id="AAG54387.1"
/ db_xref="GI:12512786"
/ translation="MISRTALSKVKGSMGSHERHALPGVIGDDLRFKGLPLCLFI
CIITAVTVVTHHTLRTTAQREQLVLERDALDIEWRNLLEENALGDHRSVERIAT
EKLQKHVPDSQENIVQK"
3457. 5223
/ gene="ftsI"
/ note="synonym: Z0094"
3457. 5223
/ gene="ftsI"
/ function="enzyme; Cell division"
/ note="Residues 1 to 588 of 588 are 100.00 pct identical
to residues 1 to 588 of 588 from Escherichia coli K-12
Strain MG1655; B0084"
/ codon_start=1
/ transl_table=11
/ product="septum formation; penicillin-binding protein 3;
peptidoglycan synthetase"
/ protein_id="AAG54388.1"
/ db_xref="GI:12512787"

```

Alignment Scores:  
Pred. No.: 0.00239  
Score: 14.00  
Percent Similarity: 100.00%

```

/translation="MKAANKTQPKRQEEHANTISWRPALLCCGICILLALAFPLGRVAM
LQIVSDMLVREGDMSLRVQQVSTSRGMITDRSGRLAVSPVPKAIWADPKVEYHDAG
GISVGDWKALANALNIPDLQSLARINAKRGFIYLARQVNPDMADYIKKLKLPGLH
LREESRRYPSGEVTAHLIGFTNVDSQIEGVEKSFQKLTGQGERVTRKDRGRVYI
EDISSYDSQAHHNLALSIDRLQALVYRELNNVAFNKAEGSSAVLVDVNTGEVLAMA
NSPYNPNLSGTPKEMRNRTITDVEPGSTVKPVMYATALQGVVRENSVLTPIY
RINGHEIKVARYSELTLTGVLQKSKNSVGVSKALAMPSSALVDVTSRFGLGKATNIG
LVGERSGLYPQKORWSDIERATFSFGTGLMVTPLQLARVATIGSYGIYPLSITKVD
PVPYGERVPESIVRTVWHMMSVALPGGGVKAARKYIAIKTKTAKKVGPDGRVI
NKYIATGAYAPASOPREALVYVINDPQAGKYIGGAVSADVFAGIINGVLTWTNIEPD
ALTGDKNEFEVINQEGTGGRS"
5210. 6697
/ gene="murE"
/ note="synonym: Z0095"
5210. 6697
/ gene="murE"
/ function="enzyme; Cell envelop: Murein sacculus,
peptidoglycan"
/ note="Residues 1 to 495 of 495 are 98.98 pct identical to
residues 1 to 495 of 495 from Escherichia coli K-12 Strain
MG1655; B0085"
/ codon_start=1
/ transl_table=11
/ product="meso-diaminopimelate-adding enzyme"
/ protein_id="AAG54389.1"
/ db_xref="GI:12512788"
/ translation="MADRNLRLDLPWPDPAPSRALREMTLDSRVAAGDLFVAVVGH
QADGRHYIPQAIAGQVAIIAEAGEATDGEIREMHGVYIYLSQNLERLSLAGRFY
HEDSNLXXVGTGTNGKTTTQLLAQMSQLLGETSAVNGTVGTLGKVIPTENTTG
SAVDVQHELAGLVDDQATFCAMEVSSHGLVQHRVAALKFAASVFTNLSRDHLDYHGD
EHYEAAKLLYSEHHCGQAIINADDEVGRRMLKLPDAVAVSMEDHINPCHGRWLKA
TEVNYHDSGATIRFSSWGDGIEHSLGAFNVSNLLALALGLYPLADLLTAA
RLQPCVGRNEVTFEGKPTVVVDYDHTDPALEKQAARLHCAGLKWCYFGCGDQDK
GKRPDLAEFEFADVAVVVDNPRTEEPRAINDILAGMLDAGHAKVMEDRAEAVTC
AVMQAKENDVVLVAGKHEDYQIVGNQRLDYSRVTVVARLLGVIA"
6694. 8052
/ gene="murF"
/ note="synonym: Z0096"
6694. 8052
/ gene="murF"
/ function="enzyme; Cell envelop: Murein sacculus,
peptidoglycan"
/ note="Residues 1 to 452 of 452 are 99.33 pct identical to
residues 1 to 452 of 452 from Escherichia coli K-12 Strain
MG1655; B0086"
/ codon_start=1
/ transl_table=11
/ product="D-alanine:D-alanine-adding enzyme"
/ protein_id="AAG54390.1"
/ db_xref="GI:12512789"
/ translation="MISVTLSQLTDILNGELQAGDITLDVATTDTTRKLTGCLFVALK
GERDAHDADQAKAGGAGALLVSRPLDIDLPLIVKDTRLAFGLAALAAWVQQVPARY
VALTSGSGKTSVKEMTAAILSQCNLTLYTAGNLNDIGVPMTLRLTPEYDYAVIELG
AHQGEIVTVTSLTRPEALVNNLAAHLGEGFSLAGVAKAKGEFSLPENGIAIMN
ADNDNLNQSVIGSRKVRFPNNAASDFTATNHYTSHCTEETLQTPGTSVDVLLP
LPGRHNTANALAAALSNSVCATLDAIKAGLANLKAIVGRFLPQLAENQLLDDSYL
ANVGSMTAAVQVLAEMFGYRVLVVDMAELGAESEACHVQVGEAAKAGDVLRSVGK
QSHAISTAGVGVEHFAEKTALITRLKSLIAEQVITILVKGSRSAAMEEVVRLQENG
TC"
8046. 9128
/ gene="mraY"
/ note="synonym: Z0097"
8046. 9128
/ gene="mraY"
/ function="enzyme; Cell envelop: Murein sacculus,
peptidoglycan"
/ note="Residues 1 to 360 of 360 are 99.72 pct identical to
residues 1 to 360 of 360 from Escherichia coli K-12 Strain
MG1655; B0087"

```

Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 3.12% Indels: 0  
 DB: 1 Gaps: 0

US-09-701-229-2 (1-448) x AE005185 (1-12791)

Qy 157 GluLeuTyrValLeuGluLeuSerSerPheGlnLeuGluThr 170  
 |||||  
 Db 9587 GAACGTACGTACTGGAACGTGCGAGCTTCAGCTGGAAACC 9628

## RESULT 13

AE000118 21757 bp DNA linear BCT 01-DEC-2000  
 LOCUS Escherichia coli K12 MG1655 section 8 of 400 of the complete  
 DEFINITION genome.  
 ACCESSION AE000118 U00096  
 VERSION AE000118.1 GI:1786262  
 KEYWORDS  
 SOURCE Escherichia coli K12  
 ORGANISM Escherichia coli K12  
 Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 Enterobacteriaceae; Escherichia.

REFERENCE 1 (bases 1 to 21757)  
 AUTHORS Blattner,F.R., Plunkett,G. III, Bloch,C.A., Perna,N.T., Burland,V.,  
 Riley,M., Collado-Vides,J., Glasner,J.D., Rode,C.K., Mayhew,G.F.,  
 Gregor,J., Davis,N.W., Kirkpatrick,H.A., Goeden,M.A., Rose,D.J.,  
 Mau,B. and Shao,Y.

TITLE The complete genome sequence of Escherichia coli K-12

JOURNAL Science 277 (5331), 1453-1474 (1997)

MEDLINE 97426617

PUBMED 9278503

REFERENCE 2 (bases 1 to 21757)

AUTHORS Blattner,F.R.

TITLE Direct Submission

JOURNAL Submitted (16-JAN-1997) Guy Plunkett III, Laboratory of Genetics,

University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.

Email: ecollegenetics.wisc.edu Phone: 608-262-2534 Fax:

608-263-7459

REFERENCE 3 (bases 1 to 21757)

AUTHORS Blattner,F.R.

TITLE Direct Submission

JOURNAL Submitted (02-SEP-1997) Guy Plunkett III, Laboratory of Genetics,

University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.

Email: ecollegenetics.wisc.edu Phone: 608-262-2534 Fax:

608-263-7459

REFERENCE 4 (bases 1 to 21757)

AUTHORS Plunkett,G. III.

TITLE Direct Submission

JOURNAL Submitted (13-OCT-1998) Laboratory of Genetics, University of

Wisconsin, 445 Henry Mall, Madison, WI 53706, USA

This sequence was determined by the E. coli Genome Project at the

University of Wisconsin-Madison (Frederick R. Blattner, director).

Supported by NIH grants HG00301 and HG01428 (from the Human Genome

Project and NCHGR). The entire sequence was independently

determined from E. coli K12 strain MG1655. Predicted open reading

frames were determined using Genemark software, kindly supplied by

Mark Borodovsky, Georgia Institute of Technology, Atlanta, GA,

30332 [e-mail: markamber@gatech.edu]. Open reading frames that

have been correlated with genetic loci are being annotated with CG

Site Nos., unique ID nos. for the genes in the E. coli Genetic

Stock Center (CGSC) database at Yale University, kindly supplied by

Mary Berlyn. A public version of the database is accessible

(http://cgsc.biology.yale.edu). Annotation of the genome is an

ongoing task whose goal is to make the genome sequence more useful

by correlating it with other data. Comments to the authors are

appreciated. Updated information will be available at the E. coli

Genome Project's World Wide Web site

(http://www.genetics.wisc.edu). \*\*\* The E. coli K12 sequence and

its annotations are periodically updated; this is version M54. No

sequence changes. Annotation updates: updated gene identifications

and products; all new functional assignments courtesy of Monica

Riley; added promoters, protein binding sites, and repeated

sequences described in reference 1. The unique numeric identifiers

beginning with a lowercase 'b' assigned to each gene (protein- or  
 RNA-encoding) are now designated as gene synonyms instead of  
 labels. This should allow them to be searched for in Entrez as gene  
 names.

FEATURES	Location/Qualifiers
source	1..21757
	/organism="Escherichia coli K12"
	/mol_type="genomic DNA"
	/strain="K12"
	/sub_strain="MG1655"
gene	/db_xref="taxon:83333"
	/complement(77..163)
	/gene="leuL"
	/note="synonym: b0075"
	/complement(77..163)
CDS	/gene="leuL"
	/function="leader; Amino acid biosynthesis: Leucine"
	/note="f28; 100 pct identical to LPL_ECOLI SW: P09149"
	/codon_start=1
	/transl_table=11
	/product="leu operon leader peptide"
	/protein_id="AAC73186.1"
	/db_xref="GI:1786263"
promoter	/translation="MTHIVRFIGLLLLNASSLRGRRVSGIOH"
	/complement(197..226)
	/note="factor Sigma70; promoter leu; documented +1 at 83735"
protein_bind	/complement(201..221)
	/note="No predicted promoter"
	/bound_moiety="LexA predicted site"
protein_bind	/complement(201..221)
	/note="central position to predicted promoter: -0.5"
protein_bind	/bound_moiety="LexA predicted site"
	/complement(354..368)
	/note="No predicted promoter"
protein_bind	/bound_moiety="Lrp documented site"
	/complement(354..368)
	/note="central position to leuABCD promoter: -156.5"
protein_bind	/bound_moiety="Lrp documented site"
	460..474
	/note="central position to predicted promoter: -28.5"
	/bound_moiety="Lrp predicted site"
promoter	462..490
	/note="factor Sigma70; predicted +1 start at 84042"
protein_bind	/complement(519..537)
	/note="central position to predicted promoter: -319.5"
protein_bind	/bound_moiety="TyrR predicted site"
	554..565
	/note="central position to predicted promoter: 64"
	/bound_moiety="AraC predicted site"
gene	646..1767
	/gene="leuO"
	/note="synonym: b0076"
CDS	646..1767
	/gene="leuO"
	/function="putative regulator; Amino acid biosynthesis: Leucine"
	/note="o373; 100 pct identical to LEUO_ECOLI SW: P10151(290 aa) but contains 59 additional N-term aa and 24 C-term residues; 100 pct identical to PIR: S40589 but contains 24 additional C-term residues"
	/codon_start=1
	/transl_table=11
	/product="probable transcriptional activator for leuABCD operon"
	/protein_id="AAC73187.1"
	/db_xref="GI:1786264"
	/translation="MTHSTAMDSVFIRTRIFMFSEYFCFFLFYHDKSYSSQLFLC IPIRELSVTVLSMPEVOTDHPETAELS POLRMVDLNLTFVDAVMOEONTTRAA HVLGMSQPVSNVAVRLKVMFNDELFRYRGIGIQTARAFQLFGSVRAQLQVONELP GSGFEPASSERVFLHCVCSPDLSDILSIQIYNHIEQIAPNHYMKFSSLNONTHERLRY QETEFVISDEDFHRPEFTSVPLFKDMVLVSNKNHPTIKGLPLKHVDVYNQHAASVLD RFASFSPQWPTVDVKQASTAYQGMMMSVLSVSQTHLVAIAPRWLAEEFAESLELQV

protein\_bind  
1802..1816  
/note="central position to predicted promoter: -39.5"  
/bound\_moiety="Lrp predicted site"  
promoter  
1813..1839  
/note="factor Sigma70; promoter livIhp4; documented+1 at  
85394"  
1833..1847  
/note="central position to predicted promoter: -8.5"  
/bound\_moiety="Lrp predicted site"  
promoter  
1838..1866  
/note="factor Sigma70; promoter livIhp3; documented+1 at  
85430"  
1915..1929  
/note="central position to predicted promoter:73.5"  
/bound\_moiety="Lrp predicted site"  
protein\_bind  
1949..1963  
/note="central position to predicted promoter:107.5"  
/bound\_moiety="Lrp predicted site"  
promoter  
1955..1984  
/note="factor Sigma70; promoter livIhp2; documented+1 at  
85534"  
1978..1992  
/note="central position to predicted promoter:136.5"  
/bound\_moiety="Lrp predicted site"  
protein\_bind  
1995..3809  
/gene="ilvI"  
/note="synonym: b0077"  
1995..3809  
/EC\_number="4.1.3.18"  
/function="enzyme; Amino acid biosynthesis: Isoleucine,  
Valine"  
/note="o604; 98 pct identical (1 gap) to 522 residues  
from ILVI\_ECOLI SW: P00893 (566 aa)"  
gene  
/codon\_start=1  
/transl\_table=11  
/product="acetylcoenzyme A synthase III, valine sensitive,  
large subunit"  
CDS  
/protein\_id="AAC73188.1"  
/db\_xref="GI:1786265"  
/translation="MENSALNCISKIGRLMCFTHFRQTVROAMELISGAEVVRSL  
IDQGVQVPGAGALDIYDALHTVGGVSLIGQVATSLIGYDAFQECMDVGISRPVVKHSF  
VTSFGATNATGATATMDSIPLVLSGGVATSLIGYDAFQECMDVGISRPVVKHSF  
LVKQEDIPQVLLKAAFGWLAASGREGVVDLPKDLINPANKLPYWPESVSMRINPT  
TGHKGQIKRAQLQTLVAAKPVYVGGGATAGCQQLKETVEALNLPVVGSLMGLGA  
PPATHRGALGLMGHGTVEANMTHNADVIFAVGVFRDRTTNLAKVCPNATVLHID  
IDPISIKVTADIPVCDARQVLEQMLELISOESAHQPLDEIRWQQLQEWARQC  
LKYDTHSEKIKPQAVIETLRLTKGDAYVTSVSDVQCHQMFAALYPPFKRWINSGL  
GTMGFGLPALGVKMAFPEETVVCVTGDSIQMNIQELSTALQYELPVLVYVNLNRRYL  
GMVKQWQDMYSGRHSQSYMQSLPDFVRLAEAYGHVGIQISHPHELESKLSEALEQVR  
NNRLVFDVTVDGSEHYVPMQIRGGMDMMLSKTERT"  
1998..2012  
/gene="ilvI"  
/note="central position to predicted promoter:156.5"  
2017..2045  
/bound\_moiety="Lrp predicted site"  
promoter  
/gene="ilvI"  
/note="factor Sigma70; promoter livIhp1; documented+1 at  
85597"  
3812..4303  
/gene="ilvH"  
/note="synonym: b0078"  
3812..4303  
/gene="ilvH"  
/EC\_number="4.1.3.18"

## Alignment Scores:

Pred. No.:	0.00385	Length:	21757
Score:	14.00	Matches:	14
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	3.12%	Indels:	0

DB: 1 Gaps: 0

US-09-701-229-2 (1-448) x AE000118 (1-21757)

Qy 157 GluLeuTyValLeuGluLeuSerPheGlnLeuGluThr 170  
|||||  
Db 13998 GAACGTGACGTGCTGGAACACTGTCGAGCTCCAGCTGGAAC 14039  
|||||

RESULT 14  
LOCUS AE004310/c  
DEFINITION Vibrio cholerae chromosome I, section 218 of 251 of the complete chromosome.  
ACCESSION AE004310 AE0003852  
VERSION AE004310.1 GI:9656963  
KEYWORDS  
SOURCE Vibrio cholerae  
ORGANISM Vibrio cholerae  
Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;  
Vibrionaceae; Vibrio.  
REFERENCE 1 (bases 1 to 22201)  
AUTHORS Heidelberg, J.F., Eisen, J.A., Nelson, W.C., Clayton, R.A., Gwinn, M.L., Dodson, R.J., Haft, D.H., Hickey, E.K., Peterson, J.D., Umayam, L., Gill, S.R., Nelson, K.E., Read, T.D., Tettelin, H., Richardson, D., Ermolaeva, M.D., Vamathevan, J., Bass, S., Qin, H., Dragol, I., Sellers, P., McDonald, L., Utterback, T., Fleischmann, R.D., Nierman, W.C. and White, O.  
TITLE DNA sequence of both chromosomes of the cholera pathogen Vibrio cholerae  
JOURNAL Nature 406 (6795), 477-483 (2000)  
MEDLINE 20406833  
PubMed 10952301  
REFERENCE 2 (bases 1 to 22201)  
AUTHORS Heidelberg, J.F., Eisen, J.A., Nelson, W.C., Clayton, R.A., Gwinn, M.L., Dodson, R.J., Haft, D.H., Hickey, E.K., Peterson, J.D., Umayam, L.A., Gill, S.R., Nelson, K.E., Read, T.D., Tettelin, H., Richardson, D., Ermolaeva, M.D., Vamathevan, J., Bass, S., Qin, H., Dragol, I., Sellers, P., McDonald, L., Utterback, T., Fleischmann, R.D., Nierman, W.C., White, O., Salzberg, S.L., Smith, H.O., Colwell, R.R., Mekalanos, J.J., Venter, J.C. and Fraser, C.M.  
TITLE Direct Submission  
JOURNAL Submitted (14-JUN-2000) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA  
FEATURES  
source  
1..22201  
/organism="Vibrio cholerae"  
/mol\_type="genomic DNA"  
/strain="N16961"  
/serotype="O1"  
/db\_xref="taxon:666"  
/chromosome="I"  
/note="biotype: El Tor"  
complement(272..1081)  
/gene="VC2391"  
complement(272..1081)  
/gene="VC2391"  
/note="similar to GB:D10483 SP:P04036 GB:M10611 PID:145710  
PID:216458; identified by sequence similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="dihydrodipicolinate reductase"  
/protein\_id="AAF95534.1"  
/db\_xref="GI:9656964"  
/translation="MVRIATAGAAAGRMGRNLVKATHONPLSELGAGSRPSSLVGV  
IGELCGIGKGIIVLNDLEQAEQFVDIIFDTAPASTLANLALCEQHGKLVIGTGTG  
TDAOROTIEQAAKKPIVMAPNYSVGNLVFKLEKAAKVMGDCVCDLEIIEAHRHKV  
DAPSGTAIGHGEATAHAMGNQLSDVAVYAREGTCGERSNRIGTFATPAGDIEGHTA  
MEADIGERVEITHTRDMTGFANGAVKAAIWLAEQPAFTYIMDVLGUNDL"  
complement(1196..1594)  
/gene="VC2392"  
complement(1196..1594)  
/gene="VC2392"  
/note="similar to GB:D10483 SP:P08337 GB:M20791 GB:X04831



```

PID:147795; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="mutator 533"
/protein_id="AAF95533.1"
/db_xref="GI:965696"
/translation="MKRIHIVAGIIFNSDQSEIFITKRPHLHKGGFWEFFPGGKVEAG
ESREQAMVLEEEIGITVTEQAFQHFDFDYTKLSLDFEMLVTFADGQPHGREGQ
GGWKIADLANRFPPEANDPVVKQVIAQFG"
gene
1580..1747
CDS
/gene="VC2393"
/length=168
/feature="VC2393"
/feature_start=1580
/feature_end=1747
/feature_note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAF95536.1"
/db_xref="GI:965696"
/translation="MDAFQATPVFLADVNFSCIPFLKATAVLAVTSCLGKIKRVAL
ATLLENKVKLA"
gene
complement(1744..4455)
CDS
/gene="VC2394"
/length=1711
/feature="VC2394"
/feature_start=1744
/feature_end=4455
/feature_note="similar to GB:D10483 SP:P10408 GB:N20791 PID:147794
PID:216512; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="preprotein translocase, SecA subunit"
/protein_id="AAF95537.1"
/db_xref="GI:965696"
/translation="MITKLLTKVIGSRNDTRLRLRKIVKEINNVPAPFALSDEELK
AKTVFRIQENGLDOLLPEAFATVREASKRVGMRFHFDVLIGGMVLHGQIAEM
PQEGKTATLALYALNPGKVHIVTNDYLAKRDAETNPLPEELGTMGVNIPN
MPQAKKAYQADILYGTNNRGGFDYLDNMAFREDRVRQRFPAVDEVDVSIILDE
ARTPLISGPAEDSDLYIRINKLIPLOKQKDESEYRGDGHFTVDKSKQVHLTE
TQGEVEELLVNGMGEQDTLYSPANISLLHVNALRAHVLFFKNVDYIVTPDGEV
VIVDEHTQMTGRMSDGLHQAVEAKGVKIQENQTLASTFONFYRLYKLSGMT
QADTAETAFQIYGLTIVIPTRKPMVRNDPVDVYRSEAEKFAAIIDIKQRYEKG
GPVLVTYSIESKLSNALKAGIKHNVLNAKFEKAEIIVAEAGKPGCAVIATNMA
GRDIDLVGSGQAKVEKLDNPTQEQIDAIAKFWQVHDVQLQAGLHLITGERHERS
RIDNLRGSRGQAGSSRFYLSMEDILLRIFTSDRAALLQSGHDEGEAELESMLS
RSIAKRAQVRGENDIRKQLLEYDDVDVQKRVYELRDELSADDISDMAQNRD
VLNAVDEYTPQSLDMDWIDKLEDRLKNDPLPIQSLDADNKLLEYEALRERII
EQAVEYTAKEQAVPAMRNPEKSVMLQTLDTLWKEHLAAMDHLRQGHLEGRYAKN
PROEYKRSFELFELLSKSDVITVLSKRVQOOEVEERMEAOORAEAEARHAQ
AOHASADAEODESNQPVYRDERKVRNEPCPGSGKKYKQCHQIN"
4809..5282
CDS
/gene="VC2395"
/length=474
/feature="VC2395"
/feature_start=4809
/feature_end=5282
/feature_note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAF95538.1"
/db_xref="GI:965696"
/translation="NRDHRPTATDELIOASKLQIOEHAKAIIILNROLQDILPKGLK
TQVRAANYRGGLVLEASAAKMKVDYERLHLIQLBQNGFGLISIEVRVNPYLK
QSKITSEDARANRPPPISEHAHVLLAIADQASDKVKRLQSLARLAKANQDD"
complement(5386..6303)
CDS
/gene="VC2396"
/length=1017
/feature="VC2396"
/feature_start=5386
/feature_end=6303
/feature_note="similar to GB:D10483 SP:P07652 PID:145848
PID:216510 PID:40864; identified by sequence similarity;
putative"
/codon_start=1
/transl_table=11
/product="UDP-3-O-3-hydroxymyristoyl N-acetylglucosamine
6-acetylase"
/protein_id="AAF95539.1"

```

```

/db_xref="GI:965696"
/translation="MIRORTLKEIVKTTGVGLSGRKRKVTTLRPAANTGIIVRTDV
NPPVDFPADPSVRMTMLCTALVNDQGVISTVEHLNAAAGMDINAIIEVDAPPIP
INDGASFPVYLLOQAGIOTLNAPKRFIRIKKPVRIEDGDKWAEFVPFNGFRMDFEIE
FNHPAIDGDDQLRDLVDFSSQGVKEISRRATFGFMDIETLQSNLCJLGGSEDCALVL
DDYRILNEGRLDFNEFVTHKLDALDGLYMAGHAIVGFRAYKSGHGLNNOLLRAVL
ADQEAWEWATFEFEEVGSFAEPNMVLA"
complement(6401..7597)
CDS
/gene="VC2397"
/length=1196
/feature="VC2397"
/feature_start=6401
/feature_end=7597
/feature_note="similar to GB:D10483 SP:P06138 GB:X02821 PID:216509
PID:40863; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="cell division protein FtsZ"
/protein_id="AAF95540.1"
/db_xref="GI:965697"
/translation="MFEPMEMSDDAVIRKVGVGGGGNAVEHVMRESIEGVFEMFIN
TDAQALRKTSGVTIQIGNITKGLGAGANQVGRDAALAEKRIKFTLGADMFPIA
AGMGGTGTGAAPVIAEVAKELGILTVAVTKPSPFEGKRLAFAPQTEELSKHVDS
LITIPNEKLLKVLGRGILTEAFASANNVKNVQGIARLITRPGMINVDEADRVTM
SEMGHAMGSGVARGEDRAEAEANAISSPLEDIDLARGVNLNITAGLMDRLDEF
ETVGNVTKAFASDNATVIGTSLDDPMADEIRTVVATIGNEKKPDIITLVAGGKSKY
APVSPVTSRVEEAAQPLHERVEVKTPQAAAPSHSSASQSVAPKEKESGYLDIPAF
LRRQAD"
complement(7628..8890)
CDS
/gene="VC2398"
/length=1262
/feature="VC2398"
/feature_start=7628
/feature_end=8890
/feature_note="similar to GB:D10483 SP:P06137 GB:X02668 GB:M36531
GB:X02821; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="cell division protein FtsZ"
/protein_id="AAF95541.1"
/db_xref="GI:965697"
/translation="MTKTVDDNIIIVGLDIGTATVSALVGEILPDQGINIAGSSPSR
GMDKGVNDLESVKSQVRAIDQAEKSHVFLSISGRHIAIRKMGMTISDE
EVSQEDMDRAIHTAKSIKIGEQRIHLVPOFTIDYQGIKNPLGLSGVMEAGVHL
ITCHNDMANRIKAVERCGLKVEOLVYSLGASNAVITEDERELGVCVVDIGATMDI
AWTGGLRHTFVFSYAGNAVTSDFAFGTPVSDAEIKVKYGCALSELVSKDDTVN
VFSGGRPSRLQRTSEVIEPRYTELGLVNTQVDSQESLRKKEGIKHHLAAGVVL
TGGAQIEGLVACAEKRVFRNPKPLEVGLTDVYKEPYHSTAVGLLHARDSQIS
DDTEYQEPKRPSTWIGIRLNRWIKQEF"
complement(8883..9665)
CDS
/gene="VC2399"
/length=882
/feature="VC2399"
/feature_start=8883
/feature_end=9665
/feature_note="similar to GB:D10483 SP:P06136 GB:X02668 GB:X02821
PID:145725; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="cell division protein FtsQ"
/protein_id="AAF95542.1"
/db_xref="GI:965697"
/translation="MINKVLEQQRITRSPQVKHACGASFFLVLLIGLLYSTIS
WMDQRLPLSKLVLDGLDHYVSALDQVRLDRLDHIGTFMSQDINVLOESVQSIPIW
SHASIRKOMPDIKTVLTFEYQVLEALNANALDKNGTVFGDITARYNGEVKLYGPDG
TAPVLKAWRDYNPFAQLGLNLISSLVNDRRAWQIILDNGLRLEKGSLEERISRF
FLIYLQGLGNKAEQSVYIDYDITDYGAAVGVFPQELQERND"
complement(9847..11367)
CDS
/gene="VC2400"
/length=1520
/feature="VC2400"
/feature_start=9847
/feature_end=11367

```

Alignment Scores:  
Pred. No.: 0.00391  
Score: 14.00  
Percent Similarity: 100.00%  
Best Local Similarity: 100.00%  
Query Match: 3.12%  
DB: 1

Length: 22201  
Matches: 14  
Conservative: 0  
Mismatch: 0  
Indels: 0  
Gaps: 0

US-09-701-229-2 (1-448) x AE004310 (1-22201)

Oy 157 GluLeuTyValLeuGluLeuSerPheGlnLeuGluThr 170  
 Db 14439 GAACGTGTGTGCTTGAGCTATCAAGCTTCCAATTAGAACT 14398

# RESULT 15

AE008699

LOCUS

DEFINITION AE008699 22348 bp DNA linear BCT 23-APR-2003  
 Salmonella typhimurium LT2, section 7 of 220 of the complete genome.

ACCESSION AE008699 AE006458

VERSION AE008699.1 GI:16418608

KEYWORDS

SOURCE

ORGANISM

Salmonella typhimurium LT2  
 Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 Enterobacteriaceae; Salmonella.

REFERENCE

AUTHORS

McClelland, M., Sanderson, K.E., Spieth, J., Clifton, S.W.,  
 Latreille, P., Courtney, L., Portolillo, S., Ali, J., Dante, M., Du, F.,  
 Hou, S., Layman, D., Leonard, S., Nguyen, C., Scott, K., Holmes, A.,  
 Grewal, N., Mulvaney, E., Ryan, E., Sun, H., Florea, L., Miller, W.,  
 Stoneking, T., Nhan, M., Waterston, R., and Wilson, R.K.  
 Complete genome sequence of Salmonella enterica serovar Typhimurium  
 LT2

TITLE

JOURNAL- Nature 413 (6858), 852-856 (2001)  
 MEDLINE 21534948  
 PUBMED 11677609

REFERENCE 2 (bases 1 to 22348)

AUTHORS

CONSRM

JOURNAL

The Salmonella typhimurium Genome Sequencing Project

Direct Submission  
 Submitted (29-MAR-2001) Genome Sequencing Center, Department of  
 Genetics, Washington University School of Medicine, 4444 Forest  
 Park Boulevard, St. Louis, MO 63108, USA  
 COMMENT Supported by NIH grant 5U 01 AI43283

COMMENT

Coding sequences below are predicted from manually evaluated  
 computer analysis, using similarity information and the programs:  
 GLIMMER; <http://www.tigr.org/softlab/glimmer/glimmer.html> and  
 GeneMark; <http://opal.biology.gatech.edu/GeneMark/>

EC numbers were kindly provided by Junko Yaburaki and the Kyoto  
 Encyclopedia of Genes and Genomes; <http://www.genome.ad.jp/kegg/>,  
 and Pedro Romero and Peter Karp at EcoCyc;  
<http://ecocyc.PangeaSystems.com/ecocyc/>

The analyses of ribosome binding sites and promoter binding sites  
 were kindly provided by Heladia Salgado, Julio Collado-Vides and  
 Reguondb;  
[http://kinich.cifn.unam.mx:8850/db/regulondb\\_intro.frameset](http://kinich.cifn.unam.mx:8850/db/regulondb_intro.frameset)

This sequence was finished as follows unless otherwise noted: all  
 regions were double stranded, sequenced with an alternate  
 chemistries or covered by high quality data (i.e., phred quality >= 30);  
 an attempt was made to resolve all sequencing problems, such  
 as compressions and repeats; all regions were covered by sequence  
 from more than one m13 subclone.

FEATURES

source

1. .22348  
 Location/Qualifiers  
 /organism="Salmonella typhimurium LT2"  
 /mol\_type="genomic DNA"  
 /strain="LT2; SGSC 1412; ATCC 700720"  
 /db\_xref="ATCC:700720"  
 /db\_xref="taxon:99287"  
 /note="LT2"  
 complement(298. .915)  
 /gene="leuB"  
 /note="synonym: STM0110"  
 complement(298. .903)  
 /gene="leuB"  
 /EC\_number="4.2.1.33"

gene

CDS

/note="3-isopropylmalate dehydratase small subunit."  
 (SW:LEU3\_SALTY)"  
 /codon\_start=1  
 /transl\_table=11  
 /product="3-isopropylmalate isomerase (dehydratase),  
 subunit with leuc"  
 /protein\_id="AAL19074.1"  
 /db\_xref="GI:16418609"  
 /translation="MABKFTQHTGLVLPVLDAAVNDTAIPKQFLQKVRTRTGFGAHLF  
 NDWFLDEKGOQNPPEVFNPEYFGASILLIARENFSGSSREHAPALATDYGKVVII  
 APSFADIFYGNSFNQLLPVTLSDAQVDELFALYKAPNGIKFEVDELAQVYKAGDKTY  
 SFKIDFRHCHMLNGLSIGTLQHEDAIAAYENKQPAFMR"  
 complement(910. .915)  
 /gene="leuB"  
 /note="putative RBS for leuB; RegulonDB:STMS1H000490"  
 complement(914. .2327)  
 /gene="leuB"  
 /note="synonym: STM0111"  
 complement(914. .2314)  
 /gene="leuB"  
 /EC\_number="4.2.1.33"  
 /note="3-isopropylmalate dehydratase large subunit.  
 (SW:LEU2\_SALTY)"  
 /codon\_start=1  
 /transl\_table=11  
 /product="3-isopropylmalate isomerase (dehydratase),  
 subunit with leuB"  
 /protein\_id="AAL19075.1"  
 /db\_xref="GI:16418610"  
 /translation="MAKTLIEKLFDAHVFEAPNETPLLYIDRLHLYHEVTSQAFDGL  
 RAHRPVQPKGTATMDHNVSTQKDNASGEMARIQMQLIKNCHEFGVELYDLNH  
 PYQIGVHVMPGEQVTLPGMTIVCGDSHTATGAFGALAFGIQSEVHEVLAATLQK  
 GRATMKIEVTGNAAPGITAQDIKDIYLAIGKTGSAGTGHVFCGDAIRALSMGRMT  
 LCNMAIEMGAKAGLVADETTFNFKGLHAPKGRDFDEAVVEYKTLTKTDGATFDTV  
 VALRAETAPQVTGTPNGOVISVTDIIPDAPSPDVERASAKALAYNGLOQGVPL  
 TDVALDKVFIGSCTNSRIEDLRAAAEYAKGRKVPQVQALVVPFGSKVPAQAGAEGLD  
 KIFTEAGFENRLPGCSNRIELRANNDRLAPGERCASTSNRNFEGRGGRGTHLVSPAMA  
 AAAVTGHFAIRSIK"  
 complement(2317. .3422)  
 /gene="leuB"  
 /note="synonym: STM0112"  
 complement(2317. .3408)  
 /gene="leuB"  
 /EC\_number="1.1.1.85"  
 /note="3-isopropylmalate dehydrogenase. (SW:LEU3\_SALTY)"  
 /codon\_start=1  
 /transl\_table=11  
 /product="3-isopropylmalate dehydrogenase"  
 /protein\_id="AAL19076.1"  
 /db\_xref="GI:16418611"  
 /translation="MSKNYHIAVLPGDIGEPEVMAQKAVMDAVRSRFRMTTSHYD  
 VGGIADNHGHPKATVEGCEQADAILFGSVGPKWENLPSPQPERGALLPLRKHF  
 KLFNLPAKLYOGLEAFCLRADIAANGFDILCVRELGTGIYFGQPKREGSQYKEK  
 AFTEVYHFEIERIARAFESAKRRKRVTSIDKANVLOSSILWREIVNDVAKTYPD  
 VELAHYIDNATWQLKDPQDFVLLCSNFGDILSDCAMITGSMGLPSASINEQG  
 FGLYEPAGGSAPDAGKNIANPQIUSLALLRLYSLLDADAATAIEQAINRALEEGV  
 RTGDLARGAAAVSTDENDGIIRYVAEGV"  
 complement(2322. .2327)  
 /gene="leuB"  
 /note="putative RBS for leuB; RegulonDB:STMS1H000491"  
 complement(3408. .4999)  
 /gene="leuA"  
 /note="synonym: STM0113"  
 complement(3408. .4979)  
 /gene="leuA"  
 /EC\_number="4.1.3.12"  
 /note="similar to E. coli 2-isopropylmalate synthase  
 (AAC73185.1); Blastp hit to AAC73185.1 (523 aa), 92%

RBS

gene

CDS

RBS

gene

CDS

```

/db_xref="GI:16418612"
/translation="MSQQVITITLRRDGEQALQSLSAKEKLOIALALERMGVDMVE
VGFVSSPDGDFESVOTIARTIKNSRVKALARCVEKIDVAQAOLKVADAFRIHTFIAT
SPMHIATKLRSFLDVERAVVMKRVNRTDDVEFSCDAGRTPVDDIARVVEAAIN
AGARTINPDVTGYTMPPEFAGIISGLYERVPNDIKAIISVTHDDGLGIAGVNSLAAY
HAGARQVEGAMNGIERAGNCALEEVIMAIKVRKDIMNVHTNINHHIEIWRISQIVSQI
CMNPITANKAIVGSGAFAGHSSGHQDGVLRKRENYEIMPTESIGLNOIOLNITSRSGR
AAVKHMEEMGYKDYDYNMDHLYDAPLADKRGQVDFDLEALAFINKQEEPEHFR
LDYFSVQSSDIATASVKACGEIKAEANGNGPVDAYIQAINRITGYDVELVKYD
LNAKGQDQALGOVDIVVNHGRFHGVLATDIVESSAKAMVHVLNNIWRAAEVEKE
LQKKAQNKENKETV"
complement(3417..3422)
RBS
/gene="leuB"
/notes="putative RBS for leuB; RegulonDB:STMSIH000492"
complement(4994..4999)
RBS
/gene="leuA"
/notes="putative RBS for leuA; RegulonDB:STMSIH000493"
complement(5067..5298)
gene
/gene="leuL"
/notes="synonym: STM0114"
complement(5067..5153)
CDS
/gene="leuL"
/notes="leu operon leader peptide. (SW:LPL_SALTY)"
/codon_start=1
/transl_table=11
/product="leu operon leader peptide"
/protein_id="AAL19078.1"
/db_xref="GI:16418613"
/translation="MSHIVRFTGLLLNFAIVGRPVGGIQH"
complement(5162..5167)
RBS
/gene="leuL"
/notes="putative RBS for leuL; RegulonDB:STMSIH000494"
complement(5271..5279)
-10_signal
/gene="leuL"
/notes="putative -10 signal for leuL;
RegulonDB:STM1TH004472"
complement(5290..5298)
-35_signal
/gene="leuL"
/notes="putative -35 signal for leuL;
RegulonDB:STM1TH004472"
5798..6754
gene
/gene="leuO"
/notes="synonym: STM0115"
5798..5803
RBS
/gene="leuO"
/notes="putative RBS for leuO; RegulonDB:STMSIH000495"
5810..6754
CDS
/gene="leuO"
/notes="probable activator protein in leuabcd operon.
(SW:LEUO_SALTY)"
/codon_start=1
/transl_table=11
/product="putative LysR family transcriptional regulator"
/protein_id="AAL19079.1"
/db_xref="GI:16418614"
/translation="MPEVKTERPHLLDMGKPOLRMVDNLLTFDVMQEQNITRAAH
TLGMSQPAVSNARVNFDELFRVYGRGQPTARAFQFGSVRQALQVLQVNEPLPG
SGFEPTSSSRVFNLCVCSPLDNILTQIYVNRVEKAPNIHVVFKAISLNONTEHOLRYQ
EFEFVTSYEERPRPEFTSVPLFKDEMVLVSRKHPRISGPLEGDDVYNEQHAVVSLDR
YASFQPWYDTFDKQSSVAYQGMALISVLNVYSQTHVAIAPRWLAEEFASIEDLIQIL
PLPLKLNSTCYLSWHEAAGRDGKHQWMDLLVSVCKR"
6941..6953
protein_bind

```

Alignment Scores:

Pred. No.:	0.00394	Length:	22348
Score:	14.00	Matches:	14
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	3.12%	Indels:	0
DB:	1	Gaps:	0

US-09-701-229-2 (1-448) x AE008699 (1-22348)

QY 157 GluLeuTyrrValLeuGluLeuSerSerPheGlnLeuGluThr 170  
 |||||  
 Db 19097 GAATTGTACGTGCTGGAATTATCCAGTTTCCAACTGGAAACT 19138  
 Search completed: August 14, 2003, 10:38:49  
 Job time : 4461 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus.p2n model

Run on: August 14, 2003, 04:47:05 ; Search time 351 Seconds  
(without alignments)  
3445.437 Million cell updates/sec

Title: US-09-701-229-2

Perfect score: 2260  
Sequence: 1 MSLIASDHFRIVVGLKSGM.....MFKNFERGRFLPAKAVEELA 448

Scoring table:

BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 3294098

Minimum DB seq length: 25  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:

-MODEL=framet.p2n.model -DEV=xlp  
-Q/cgn2\_1/USPTO.spool/US09701229/runat\_09082003\_161156\_4371/app\_query.fasta\_1.647  
-DB=N\_Geneseq\_19Jun03 -Qfmt=fastp -SUFFIX=p2n.rng -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blom62 -TRANS=human40.cdi  
-LIST=45 -DOCALIGN=200 -THR\_SCORE=pcpt -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTPMT=pto -NORM=ext -HEAPSIZ=500 -MINLEN=25 -MAXLEN=2000000000  
-USER=US09701229.ecgn.1.1.471/runat\_09082003\_161156\_4371 -NCPU=6 -ICPU=3  
-NO\_MAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSPBLOC=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N\_Geneseq\_19Jun03:\*\*

1: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1980.DAT:\*\*  
2: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1981.DAT:\*\*  
3: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1982.DAT:\*\*  
4: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1983.DAT:\*\*  
5: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1984.DAT:\*\*  
6: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1985.DAT:\*\*  
7: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1986.DAT:\*\*  
8: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1987.DAT:\*\*  
9: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1988.DAT:\*\*  
10: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1989.DAT:\*\*  
11: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1990.DAT:\*\*  
12: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1991.DAT:\*\*  
13: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1992.DAT:\*\*  
14: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1993.DAT:\*\*  
15: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1994.DAT:\*\*  
16: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1995.DAT:\*\*  
17: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1996.DAT:\*\*  
18: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1997.DAT:\*\*  
19: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1998.DAT:\*\*  
20: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1999.DAT:\*\*  
21: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2000.DAT:\*\*  
22: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT:\*\*  
23: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT:\*\*  
24: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2002.DAT:\*\*  
25: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2003.DAT:\*\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed,

and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
1	2260	100.0	1347	23	Pseudomonas aerugi
2	2260	100.0	1450	21	Pseudomonas aerugi
3	989.5	43.8	1317	23	Salmonella typhi D
4	985.5	43.6	1317	22	Escherichia coli p
5	985.5	43.6	1317	23	E. coli DNA for ce
6	889.5	39.4	96109	22	Genomic fragment #
c	879	38.9	1314	23	Haemophilus influe
7	879	38.9	1314	23	Haemophilus influe
8	879	38.9	1314	23	Haemophilus influe
9	716	31.7	640681	24	Buchnera sp. genom
c	704	31.2	1380	21	Neisseria meningit
10	704	31.2	1380	21	Neisseria meningit
c	704	31.2	1380	21	Neisseria meningit
11	704	31.2	1380	21	Neisseria meningit
12	704	31.2	1380	21	Neisseria meningit
13	704	31.2	1380	21	Neisseria meningit
14	701	31.0	1335	25	N. gonorrhoeae nuc
15	603.5	26.7	2193	20	Streptococcus pneu
16	594.5	26.3	1350	25	S. pneumoniae type
c	594.5	26.3	1350	25	S. pneumoniae type
17	594.5	26.3	1350	25	S. pneumoniae type
18	594.5	26.3	1350	25	S. pneumoniae type
19	592.5	26.2	1353	23	Streptococcus pneu
20	586.5	26.0	1350	18	Streptococcus pneu
21	586.5	26.0	1350	20	S. pneumoniae MURD
22	586.5	26.0	1353	23	S. pneumoniae MURD
23	586.5	26.0	1353	23	Streptococcus pneu
24	577.5	25.6	6217	19	DNA encoding novel
25	542	24.0	1080	23	DNA encoding novel
26	542	24.0	1080	23	DNA encoding novel
27	542	24.0	1080	23	DNA encoding novel
28	537.5	23.8	2703	23	DNA encoding novel
29	537.5	23.8	2703	23	DNA encoding novel
30	534.5	23.7	1352	24	Bacillus lichenifo
31	529.5	23.4	1353	24	Streptococcus poly
32	529.5	23.4	1365	24	Streptococcus poly
33	529.5	23.4	1365	24	Streptococcus poly
34	525	23.2	215561	22	Streptococcus poly
c	525	23.2	215561	22	Streptococcus poly
35	525	23.2	215561	22	Streptococcus poly
36	519	23.0	1368	23	Mycobacterium tube
37	519	23.0	1371	23	Enterococcus faeca
c	515	22.8	4403765	22	Enterococcus faeca
38	515	22.8	4403765	22	Enterococcus faeca
39	510	22.6	4942	23	Mycobacterium tube
40	510	22.6	4944	23	DNA encoding novel
41	510	22.6	4944	23	DNA encoding novel
42	509.5	22.5	2365589	24	DNA encoding novel
c	506.5	22.4	1356	24	Genomic sequence o
43	506.5	22.4	1356	24	Streptococcus poly
44	506.5	22.4	1407	20	Streptococcus pyog
45	493	21.8	1003	19	Streptococcus pneu

ALIGNMENTS

RESULT 1

AAS54275  
ID AAS54275 standard; DNA; 1347 BP.

AC AAS54275;

DT 13-FEB-2002 (first entry)

XX Pseudomonas aeruginosa DNA for cellular proliferation protein #406.

XX Antisense; ds; prokaryotic cellular proliferation gene;

KW antibiotic; antibacterial; drug design.

XX Pseudomonas aeruginosa.

XX WO200170955-A2.

XX

PD 27-SEP-2001.  
 PF 21-MAR-2001; 2001WO-US09180.  
 XX 21-MAR-2000; 2000US-191078P.  
 PR 23-MAY-2000; 2000US-206948P.  
 PR 26-MAY-2000; 2000US-207727P.  
 PR 23-OCT-2000; 2000US-242578P.  
 PR 27-NOV-2000; 2000US-253625P.  
 PR 22-DEC-2000; 2000US-257931P.  
 PR 16-FEB-2001; 2001US-269308P.  
 XX (ELIT-) ELITRA PHARM INC.  
 PA Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;  
 PI Yamamoto RT, Xu HH;  
 XX WPI: 2001-611495/70.  
 DR P-PSDB; A036416.  
 XX  
 PT New polynucleotides for the identification and development of  
 PT antibiotics, comprise sequences of antisense nucleic acids -  
 PS Claim 27; Seq ID No 7912; 511pp; English.  
 XX  
 CC The invention relates to antisense inhibitors of genes essential to  
 CC prokaryotic cellular proliferation, their use in identifying the  
 CC genes, their use in the discovery of novel antibiotics, the essential  
 CC genes themselves and the encoded proteins. The prokaryotes used are  
 CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella  
 CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The  
 CC invention is also useful for the identification of potential new targets  
 CC for antibiotic development. The antisense nucleic acids can also be used  
 CC to identify proteins used in proliferation, to express these proteins,  
 CC and to obtain antibodies capable of binding to the expressed proteins.  
 CC The proteins can be used to screen compounds in rational drug discovery  
 CC programmes. The antisense nucleic acid sequence is also useful to screen  
 CC for homologous nucleic acids which are required for cell proliferation in  
 CC a wide variety of organisms. The present sequence encodes an  
 CC essential prokaryotic cellular proliferation protein.  
 CC Note: The sequence data for this patent did not form part  
 CC of the printed specification, but was obtained in electronic  
 CC format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 1347 BP; 214 A; 453 C; 453 G; 227 T; 0 other;

Alignment Scores:  
 Pred. No.: 1,27e-205 Length: 1347  
 Score: 2260.00 Matches: 448  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 DB: 23 Gaps: 0

US-09-701-229-2 (1-448) x AAS54275 (1-1347)

QY 1 MetSerLeuAlaSerAspHisPheArgIleValValGlyLeuGlyLysSerGlyMet 20  
 DB 1 ATGAGCCGTGATCGCTCCGACCACCTTCGCGCATCGTTGTCGGCTCGGCAAGCGGCATG 60  
 QY 21 SerLeuValArgTyrLeuAlaArgArgGlyLeuProPheAlaValValAspThrArgGlu 40  
 DB 61 TCCCTGGTGGCGTACCTGGCGCGCGCGCTTTCGCGCTTTCGCGCTGATACCGAGAG 120  
 QY 41 AsnProGluLeuAlaThrLeuArgAlaGlnTyrProGlnValGluValArgCysGly 60  
 DB 121 AACCCGCGGAGCTGGCCACCCCTTCGCTGCCAGATATCCGAGGTGAAGTGGCTTGGCGC 180  
 QY 61 GluLeuAspAlaGluPheLeuCysSerAlaArgGluLeuTyrValSerProGlyLeuSer 80  
 DB 181 GAACCTCGAGCGGAGTTCCTCTGTCTCCCGCGGACACTCTATGTCAGCCCGCGGCTTGTG 240

QY 81 LeuArgThrProAlaLeuValGlnAlaAlaLysGlyValArgIleSerGlyAspIle 100  
 DB 241 CTGCGACCCCTGCGCTGATACAGGCGCGGAAAGGCGTGCATCTCCGGTGACATC 300  
 QY 101 AspLeuPheAlaArgGluAlaLysAlaProIleValAlaIleThrGlySerAsnAlaLys 120  
 DB 301 GATCTCTTCGCGCGGAGCGGAAAGCGCGCATGTCGCCCATCACCGTTTCCAAACGGAAG 360  
 QY 121 SerThrValThrThrLeuValGlyGluMetAlaValAlaAlaAspLysArgValAlaVal 140  
 DB 361 AGCACCGTGACCCCTGCTGGCGGAAATGGCGGTGGCGCGGACAAAGCGTCTGCCCGTC 420  
 QY 141 GlyGlyAsnLeuGlyThrProAlaLeuAspLeuLeuAlaAspIleGluLeuTyrVal 160  
 DB 421 GCGCGCAACCTCGGCGACCCCGCGCTCGACCTGCTGGCGGAGACATCGAGCTGACGTG 480  
 QY 161 LeuGluLeuSerSerPheGlnLeuGluThrCysAspArgLeuAsnAlaGluValAlaThr 180  
 DB 481 TTGGAGCTGTGAGCTTCCAGCTGGAACCTCGCATCGCTCAAGCCGAGGTGGCGACC 540  
 QY 181 ValLeuAsnValSerGluAspHisMetAspArgTyrAspGlyMetAlaAspTyrHisLeu 200  
 DB 541 GTGCTGAACGTCAGCGAAGACCATATGGATCGCTACGACGGCATGGCTGACTACACCTG 600  
 QY 201 AlaLysHisArgIlePheArgGlyAlaArgGlnValValValAsnArgAlaAspAlaLeu 220  
 DB 601 GCCAAGCACCGGATCTTCGCGGTGCCGCCAGGTGCTGGTGAATCGCGCCGATGCCCTG 660  
 QY 221 ThrArgProLeuIleAlaAspThrValProCysTyrTrpSerPheGlyLeuAsnLysProAsp 240  
 DB 661 ACCCGACCGGTATCGCGATACCGTGGCTGCTGGTGGCTGCTGGCTGCTGGCTGCTGGCTG 720  
 QY 241 PheLysAlaPheGlyLeuIleGluAspGlyGlnLysTrpLeuAlaPheGlnPheAsp 260  
 DB 721 TTCAAGGCTTTCGCGCTGATCGAGGAAGACGCCAAGAGTGGCTGGCTTCCAGTTCGAC 780  
 QY 261 LysLeuLeuProValGlyLeuLysIleArgGlyAlaHisAsnTyrSerAsnAlaLeu 280  
 DB 781 AAGTGTCTGCGGTTGGCGAATCGATGCGTGGCGGCCCAACATATTTCACACGCGTTC 840  
 QY 281 AlaAlaLeuAlaLeuGlyHisAlaValGlyLeuProPheAspAlaMetLeuGlyAlaLeu 300  
 DB 841 GCCGCGCTGGCGTGGCGCATCGCGTGGCTGCGCTGCGCTGCGAGCCCATGCTCGCGCGCTG 900  
 QY 301 LysAlaPheSerGlyLeuAlaHisArgCysGlnTrpValArgGluArgGlnGlyValSer 320  
 DB 901 AAGCGCTTTCGCGCTGCTCATCGCTGCCAGTGGGTACGCGAGCGGCGGCGCGCGGAT 960  
 QY 321 TyrTyrAspAspSerLysAlaThrAsnValGlyAlaAlaLeuAlaAlaIleGluGlyLeu 340  
 DB 961 TACTACGACGATTCCAAAGGCCACCAACGTCGCGCGCGCGCTGGCGCGCATCGAGGGGCTG 1020  
 QY 341 GlyAlaAspIleAspGlyLysLeuValLeuLeuAlaGlyLysGlyAspGlyLysGlyAlaAsp 360  
 DB 1021 GGTGCCGACATCGACGCGAAGCTGTGTGCTGCGCGCGGAGAGAGCGGCGGCGCGGAT 1080  
 QY 361 PheHisAspLeuArgGluProValAlaAlaArgPheCysArgAlaValValLeuLeuGlyArg 380  
 DB 1081 TTCCATGACCTGCGGAGCGGCTGCGCGCTTCTCCGCGCGGTGGTACTGTTGGCGCT 1140  
 QY 381 AspAlaGlyLeuIleAlaGlnAlaLeuGlyAsnAlaValProLeuValArgValAlaThr 400  
 DB 1141 CACGCGGCGGTGATTGGCCAGCGACTGGGCGAAGCGGTGACCGCTGGTGGCTGCGCTGCAAG 1200  
 QY 401 LeuAspGluAlaValArgGlnAlaAlaGluLeuAlaArgGluGlyAspAlaValLeuLeu 420  
 DB 1201 CTGGACGAAGCAGTCCGCGAGCGCCCGAGCTGGCGCGCGGAGAGCGGCGGTGCTGTG 1260  
 QY 421 SerProAlaCysAlaSerLeuAspMetPheLysAsnPheGluGluArgGlyArgLeuPhe 440  
 DB 1261 TCGCGCGCTCGCGGAGCTGGACATGTTCAAGAACTTCGAAGAACTCGAAGACGCGGCGCTGTC 1320  
 QY 441 AlaLysAlaValGluGluLeuAla 448

```

Db      1321 GCCAAAGCCGTAGAGGAGTAGCG 1344
RESULT 2
AAZ47131
ID      AAZ47131 standard; DNA; 1450 BP.
XX      AC
XX      AAZ47131;
XX      28-MAR-2000 (first entry)
XX      Pseudomonas aeruginosa murD gene.
DE      MurD; bacterial cell wall; biosynthesis; inhibitor; D-glutamate;
KW      UDP-N-acetylmuramyl-L-alanine precursor; ds.
XX      Pseudomonas aeruginosa.
OS      W09961050-Al.
FN      02-DEC-1999.
PD      26-MAY-1999; 99WO-US11585.
PF      29-MAY-1998; 98US-0087308.
PR      (MERI ) MERCK & CO INC.
XX      El-Sherbelini M, Azzolina B;
PI      WPI; 2000-072548/06.
DR      P-PSDB; AAY56047.
XX      New nucleic acid encoding the MurD protein of Pseudomonas aeruginosa,
PT      used to identify specific inhibitors -
XX      Claim 4; Page 8-9; 35pp; English.
XX      This sequence represents the MurD gene from Pseudomonas aeruginosa.
CC      MurD is an enzyme involved in biosynthesis of bacterial cell walls,
CC      so inhibiting it inhibits bacterial growth (by preventing
CC      MurD-catalyzed addition of D-glutamate to an alanyl residue in the
CC      UDP-N-acetylmuramyl-L-alanine precursor). Cells that express murD
CC      protein can be used: (a) to identify inhibitors of MurD, and (b) for
CC      recombinant production of MurD, which is used to raise specific
CC      antibodies (potential therapeutic inhibitors). Fragments of the murD
CC      gene can be used as probes and primers to detect MurD-encoding sequences
CC      (to identify polymorphic murD genes or to detect P. aeruginosa), also
CC      as antisense modulators. The murD inhibitors are useful for treating or
CC      preventing bacterial infections, including those caused by P. aeruginosa.
XX      Sequence 1450 BP; 227 A; 487 C; 487 G; 249 T; 0 other;

Alignment Scores:
Pred. No.:      1.4e-205      Length:      1450
Score:          2260.00      Matches:      448
Percent Similarity: 100.00%      Conservative: 0
Best Local Similarity: 100.00%      Mismatches: 0
Query Match:    100.00%      Indels:      0
DB:            21      Gaps:        0

US-09-701-229-2 (1-448) x AAZ47131 (1-1450)
Qy      1 MetSerLeuIleAlaSerAspHisPheArgTleValValGlyLeuGlyLysSerGlyMet 20
Db      51 ATGAGCCTGATCGCTCCGACCACCTCCGGATCGTTCGGCTCGGCAAGAGCGGCATG 110
Qy      21 SerLeuValArgTyrLeuAlaArgArgGlyLeuProPheAlaValAlaAspThrArgGlu 40
Db      111 TCCCTGTGGCTACCTTGGCGCGCGCGCTTTCGCTTCGCGGTGTCGATACCCGAGAG 170
Qy      41 AsnProProGluLeuAlaThrLeuArgAlaGlnTyrProGlnValGluValArgCysGly 60

```

QY 421 SerProAlaCysAlaSerLeuaspMetPheLysAsnPheGluGluArgGlyArgLeuPhe 440  
 |||||  
 Db 1311 TCGCCGCGCTGGCGAGCTGGACATGTTCAAGAACTTCAAGAGCGCGAGCGCTGTTC 1370  
 |||||  
 QY 441 AlaLysAlaValGluGluLeuAla 448  
 |||||  
 Db 1371 GCCAAGCGTAGAGGAGCTAGCG 1394

## RESULT 3

AAS56341

ID AAS56341 standard; DNA; 1317 BP.

AC AAS56341;

DT 13-FEB-2002 (first entry)

DE Salmonella typhi DNA for cellular proliferation protein #374.

KW Antisense; ds; prokaryotic cellular proliferation gene;

XX antibiotic; antibacterial; drug design.

OS Salmonella typhi.

PN WO200170955-A2.

XX 27-SEP-2001.

XX 21-MAR-2001; 2001WO-US09180.

XX 21-MAR-2000; 2000US-191078P.

XX 23-MAY-2000; 2000US-206848P.

XX 26-MAY-2000; 2000US-207272P.

XX 23-OCT-2000; 2000US-242578P.

XX 27-NOV-2000; 2000US-253625P.

XX 22-DEC-2000; 2000US-257931P.

XX 16-FEB-2001; 2001US-269308P.

XX (ELIT-) ELITRA PHARM INC.

PA Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;

XX Yamamoto RT, Xu HH;

PI WPI: 2001-611495/70.

DR P-PSDB; AAU38482.

XX New polynucleotides for the identification and development of

PT antibiotics, comprise sequences of antisense nucleic acids -

XX Claim 27; Seq ID No 9978; 51lpp: English.

XX The invention relates to antisense inhibitors of genes essential to

CC prokaryotic cellular proliferation, their use in identifying the

CC genes, their use in the discovery of novel antibiotics, the essential

CC genes themselves and the encoded proteins. The prokaryotes used are

CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella

CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The

CC invention is also useful for the identification of potential new targets

CC to identify proteins used in proliferation, to express these proteins,

CC and to obtain antibodies capable of binding to the expressed proteins.

CC The proteins can be used to screen compounds in rational drug discovery

CC programmes. The antisense nucleic acid sequence is also useful to screen

CC for homologous nucleic acids which are required for cell proliferation in

CC a wide variety of organisms. The present sequence encodes an

CC essential prokaryotic cellular proliferation protein.

CC Note: The sequence data for this patent did not form part

CC of the printed specification, but was obtained in electronic

CC format directly from WIPO at

CC ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 1317 BP; 258 A; 341 C; 427 G; 291 T; 0 other;

## Alignment Scores:

Pred. No.: 1,24e-84 Length: 1317  
 Score: 989.50 Matches: 230  
 Percent Similarity: 64.56% Conservatives: 56  
 Best Local Similarity: 51.92% Mismatches: 136  
 Query Match: 43.78% Indels: 21  
 DB: 23 Gaps: 9

US-09-701-229-2 (1-448) x AAS56341 (1-1317)

QY 11 IleValValGlyLeuLysSerGlyMetSerLeuValArgTyrLeuAlaArgArgGly 30  
 |||||  
 Db 28 GTCATTATCGGTCTGGCTTACCGGACTCTCTGTGGTGGACTTTTCTCGCCCGCGGC 87  
 |||||  
 QY 31 LeuProPheAlaValValAspThrArgGluAsnProGluLeuAlaThrLeuArgAla 50  
 |||||  
 Db 88 GTGACGCCGCGGTGATGATGATCTCGCTGACGCCGCGGTCTGTGTAAGCTG----- 141  
 |||||  
 QY 51 GlnTyrProGln---ValGluValArgCysGlyGluLeuAspAlaGluPheLeuCysSer 69  
 |||||  
 Db 142 -----CCGCAAGAGGTTGAGGCTCAGTTGGCGGCTGAACGACGAGTGGCTTATGCG 195  
 |||||  
 QY 70 AlaArgGluLeuTyrValSerProGlyLeuSerLeuArgThrProAlaLeuGlnAla 89  
 |||||  
 Db 196 GCGGATTATCGTCCGACCCCTGGTATTGCTGCGGCATCCCTCTCTTACCGCTGCC 255  
 |||||  
 QY 90 AlaAlaLysGlyValArgIleSerGlyAspIleAspLeuPheAlaArgGluAlaLysAla 109  
 |||||  
 Db 256 GCCAGCGCTGGAGTGGAAATTTGCGCGATATCGAACTGTTTGGCGCGAAGCGAAGCG 315  
 |||||  
 QY 110 ProIleValAlaIleThrGlySerAsnAlaLysSerThrValThrThrLeuValGlyGlu 129  
 |||||  
 Db 316 CCGATTGTGGCCATCACCAGCTCGAAGCAAAAGCAGCCGTGACCACTTAGTGGCGAG 375  
 |||||  
 QY 130 MetAlaValAlaAlaAspLysArgValAlaValAlaGlyGlyAsnLeuGlyThrProAlaLeu 149  
 |||||  
 Db 376 ATGGCAAGCGCGCGGCGCTCAATGCGCGTGGCGCAATCGTCTTCCGCGCGCTG 435  
 |||||  
 QY 150 AspLeuLeuAlaAspAspIleGluLeuTyrValLeuGluLeuSerSerPheGlnLeuGlu 169  
 |||||  
 Db 436 ATGCTGCTGGATGCGCGATCGCGAATTTGTACGTGTGGAATTTATCCAGTTTCCAAATGGAA 495  
 |||||  
 QY 170 ThrCysAspArgLeuAsnAlaGluValAlaThrValLeuAsnValSerGluAspHisMet 189  
 |||||  
 Db 496 ACTACCTCAAGTTTTCAGCGCGCGCGCAAGCGTGTCAAGCTCACTGAAGATCATATG 555  
 |||||  
 QY 190 AspArgTyrAsp---GlyMetAlaAspTyrHisLeuAlaLysHisArgIlePheArgGly 208  
 |||||  
 Db 556 GACCGGTATCGTGTGGTTTGGCAACAGTACCGCGCGCGCAAACTGCGCGCTCTACGAGAAG 615  
 |||||  
 QY 209 AlaArgGlnValValAlaAsnArgAlaAspAlaLeuThrArgProLeuIleAlaAspThr 228  
 |||||  
 Db 616 GCGAAAGTGTGGTAGTGAATCCCGATGACGCGTGTGACGATCGCGTACGCGGGCGGAT 675  
 |||||  
 QY 229 ValProCysTrpSerPheGlyLeuAsnLysProAspPheLysAlaPheGlyLeuIleGlu 248  
 |||||  
 Db 676 GAGCGCTCGCTCAGCTTTGGCGTCAATATGGGTGATATATCAC-----CTTAATCGT 726  
 |||||  
 QY 249 GluAspGlyGlnLysTrpLeuAlaPheGlnPheAspLysLeuLeuProValGlyGluLeu 268  
 |||||  
 Db 727 CAGCAGCGGGAACCTGGCTAGCTCAAGGTGAGAGGTGCTCAACGTGAAGAGATG 786  
 |||||  
 QY 269 LysIleArgGlyAlaHisAsnTyrSerAsnAlaLeuAlaLeuAlaLeuGlyHisAla 288  
 |||||  
 Db 787 AAGCTTTCGGTCAAGTAATATACCAATGTTAGCGCGCTGCGCGTGGCGGATGCC 846  
 |||||  
 QY 289 ValGlyLeuProPheAspAlaMetLeuGlyAlaLeuLysAlaPheSerGlyLeuAlaHis 308  
 |||||  
 Db 847 GTAGTCTCGCGCGCGCGAGTGTGAAGCGTGTGACGACATTTACCGGTCTGCGCAT 906  
 |||||  
 QY 309 ArgCysGlnTrpValArgGluArgGlnGlyValSerTyrTyrAspSerLysAlaThr 328  
 |||||  
 Db 907 CGCTTCCAGCTGGCGTTGGAGCATAAACGCGCTTCGTTGGATCAACGACTCGAAGACGACC 966  
 |||||



Qy	329	AsnValGlyAlaAlaLeuAlaLalleGlugLyLeuGlyAlaAspIleAspGlyLysLeu	348
		:::	:::
Db	967	AATGTCCGCAGTACCGAAGCGGCTAAACGGTTTG-----CATGTGGACGGTAGCCTG	1020
		:::	:::
Qy	349	valLeuAlaClglyaspGlyLysGlyAlaAspPheHisaspLeuArgGluProVal	368
		:::	:::
Db	1021	CATCTGCTGTGGCGCGCACGGTAAGTCGGCAGACTTT-----TCTCGGCTG	1068
		:::	:::
Qy	369	AlaArgPhecysArgAla-----ValValLeuGlyArgaspAlaGlyLeu	384
		:::	:::
Db	1069	GCGCGTATTGTACC GGCGCATGCTATCCGCTGTATTGCTTTGGGCGCGCATGGCGCGCAG	1128
		:::	:::
Qy	385	IleAlaGlnAlaLeuGlyAsnAlaValProLeuValArgValAlaThrLeuAspGluAla	404
		:::     :::	:::     :::
Db	1129	CTTGCC---GCATCGGTCCGGGAATC---GCCCAACAGACTGAGAGTGGGAAGAGCGG	1182
		:::	:::
Qy	405	ValargGlnAlaAlaGluLeuAlaArgGluGlyAspAlaValLeuLeuSerProAlaCys	424
		:::     :::	:::     :::
Db	1183	ATCGGTTTGTGTGCGCGCGCACGTTTCAGCCGGGTGATATGGTGTGTGTGTGCGCCGCTGC	1242
		:::	:::
Qy	425	AlaSerLeuAspMetPhelyAsnPheGluGluArgLeuPheAlaLysAlaVal	444
		:::	:::
Db	1243	GCCAGCCTCGATCAGTTTAAAAATTTTGAGCAACGGGCGATGCTCTTACCCTGCTCGCG	1302
		:::	:::
Qy	445	GluGluLeu	477
		:::	:::
Db	1303	AAGCAGTTA	1311

## RESULT 4

AAH81460	AAH81460 standard; DNA; 1317 BP.
XX	AAH81460;
XX	AAH81460;
XX	21-SEP-2001 (first entry)
DT	
XX	Escherichia coli protein encoding nucleotide sequence SEQ ID NO:259.
XX	
DE	Escherichia coli; identification; proliferation; microorganism;
XX	antimicrobial; antibacterial; antibiotic; gene therapy; diagnosis;
KW	bacterial growth inhibition; ds.
KW	
XX	
XX	Escherichia coli.
OS	
XX	WO200148209-A2.
XX	
PN	05-JUL-2001.
XX	
PD	
XX	19-DEC-2000; 2000WO-US34419.
XX	
PF	
XX	23-DEC-1999; 99US-0173005.
PR	
XX	(ELIT-) ELITRA PHARM INC.
XX	
PA	
XX	Forsyth RA, Ohlsen KL, Zyskind JW;
XX	
PI	
XX	WPI; 2001-457376/49.
DR	
DR	P-PSDB; AAG98404.
DR	
XX	
XX	Novel nucleic acids encoding proteins required for Escherichia coli
PT	proliferation, useful for screening for antimicrobial agents -
PT	
XX	Claim 9; Page 388-389; 596pp; English.
PS	
XX	
CC	The present invention describes a purified or isolated nucleic acid
CC	sequence (I) consisting essentially of one of the 93 nucleotide sequences
CC	given in AAH81202 to AAH81294, where expression of the nucleic acid in a
CC	microorganism is capable of inhibiting proliferation of a microorganism.
CC	(I) have antibacterial and antibiotic activities, and can be used in
CC	gene therapy. Expression of (I) in a microorganism inhibits proliferation
CC	of the microorganism, and the manufactured antibiotic is useful for
CC	reducing the activity or level of a gene product required for

proliferation of a microorganism in a subject, specifically humans. The nucleic acids that inhibit bacterial growth or proliferation can be used as antisense therapeutics for killing bacteria. In addition to therapeutic applications, the nucleic acid sequences complementary to sequences required for proliferation can be used as diagnostic tools. For example, nucleic acid probes complementary to proliferation-required sequences that are specific for particular species of microorganisms can be used as probes to identify particular microorganism species in clinical specimens. AAH81295 to AAH81487 encode the *Escherichia coli* proteins given in AAG98239 to AAG98431, and AAH91488 to AAH91491 represent oligonucleotides, which are used in the exemplification of the present invention.

Sequence 1317 BP: 294 A: 323 C: 400 G: 300 T: 0 other;

Alignment Scores:	
Pred. NO.:	2,988-84
Score:	985.50
Percent Similarity:	64.3%
Best Local Similarity:	51.24%
Query Match:	43.61%
DB:	22
Length:	1317
Matches:	227
Conservative:	58
Mismatches:	137
Indels:	21
Gaps:	9

US-09-701-229-2 (1-448) x AAH81460 (1-1317)

Qy	11	IleValValGlyLeuGlyLysSerGlyMetSerLeuValArgTyrLeuAlaArgGly	30
Db	28	GTCAATATACGGCTGGGGCTCACCGGGCTTTCCTCGCTGGACTTTTCTCCTCGCGGT	87
Qy	31	LeuProPheAlaValAlaSerThrArgGluAsnProGluLeuAlaThrLeuArgAla	50
Db	88	GTGACGGCGCGGTATGGATACGGGTATGACACCGCTGGCTGGATAAAATTA	141
Qy	51	GlnTyrProGln---ValGluValArgCysGlyGluLeuAspAlaGluPheLeuCysSer	69
Db	142	-----CCCGAAGCCGTAGAACCCACAGGGCAGCTCTGAATGATGAATGGCTGATGGC	195
Qy	70	AlaArgGluLeuTyrValSerProGlyLeuSerLeuArgThrProAlaLeuValGlnAla	89
Db	196	GCAGATCTGATGTTCGCCAGTCCCGGTATTGCACATGGCGCATCCATCTTAAGCGCTGC	255
Qy	90	AlaAlaLysGlyValArgIleSerGlyAspIleAspLeuPheAlaArgGluAlaLysAla	109
Db	256	GCTGATGCCGAATCGAATCGTTGGCGATATCGAGCTGTTCTGTGGCGAAGCACAAGCA	315
Qy	110	ProIleValAlaIleThrGlySerAsnAlaLysSerThrValThrThrLeuValGlu	129
Db	316	CCGATTGGCGGATATACCGGTTCTAACCGCAAAACACGGCTCACCGCTAGTGGGTGAA	375
Qy	130	MetAlaValAlaAlaAspLysArgValAlaValGlyGlyAsnLeuGlyThrProAlaLeu	149
Db	376	ATGCGAAAGCGGGGGGTTAAACGTTGGTGGGTGGCAATATTGGCTGCCTGCGTTG	435
Qy	150	AspLeuLeuAlaAspAspIleGluLeuTyrValLeuGluLeuSerSerPheGlnLeuGlu	169
Db	436	ATGCTACTGGATGATGAGTGTAACGTGACGTGCGTGGAACTGTGAGCTTCAGCTGGAA	495
Qy	170	ThrCysAspArgLeuAsnAlaGluValAlaThrValLeuAsnValSerGluAspHisMet	189
Db	496	ACCACCTCCAGCTTACAGGGGTACAGCGACCATCTGAACGTGATGAAGATCATATG	555
Qy	190	AspArgTyrAsp---GlyMetAlaAspTyrHisLeuAlaLysHisArgIlePheArgGly	208
Db	556	GATCGTATCGTTGGTTTACAAACAGTATCGTCGAGCAAAACTGGCATTTACGAAAC	615
Qy	209	AlaArgGlnValValAlaAsnArgAlaAspAlaLeuThrArgProLeuIleAlaAspThr	228
Db	616	GCGAAGTTTCGTGGTTAATGCTGATGATGCTTTAAACAATCCGATTCCGCGTGGCGAT	675
Qy	229	ValProCysTrpSerPheGlyLeuAsnLysProAspPheLysAlaPheGlyLeuIleGlu	248
Db	676	GAACGCTCGCTCAGCTTGGGCTCAACATGGGTGACTATCAC-----CTGAATCAT	726

D**b**



150 AspLeuLeuAlaAspAspIleGluLeuTyrrValLeuGluLeuSerSerPheGlnLeuGlu 169  
 436 ATGCTACTGGATGATGAGTGTGAAGCTGACCTGCTGGAACATGTCGAGCTCCAGCTGGAA 495  
 170 ThrCysAspArgLeuAsnAlaGluValAlaThrValLeuAsnValSerGluAspHisMet 189  
 496 ACCACCTCCAGCTTACAGCGCGTATGACGCCGACCATCTGAACGTGACTGAAGATCATATG 555  
 190 AspArgTyrAsp---GlyMetAlaAspTyrHisLeuAlaLysHisArgIlePheArgGly 208  
 556 GATCGCTATCCGTTTGGTTTACAACAGTATCGTCGACAAACATGCGCATTTACGAAAC 615  
 209 AlaArgGlnValValValasnArgAlaAspAlaLeuThrArgProLeuIleAlaAspThr 228  
 616 GCGAAAGTTTCCGCGTGTAAATGCTGATGATGCCCTTACAAATGCGGATTCGCGTGGCGGAT 675  
 229 ValProCysTyrSerPheGlyLeuAsnLysProAspPheLysAlaPheGlyLeuIleGlu 248  
 676 GAACGCTGCGTCAGCTTTGGCGTCAACATCGGTGACTATCAC-----CTGAATCAT 726  
 249 GluAspGlyGlnLysTyrLeuAlaPheGlnPheAspLysLeuLeuProValGlyGluLeu 268  
 727 CAGCAGCGCGAAACCTCGCTGCGGGTTAAAGCGGAGAAAGTCTGTAATGTGAAGAGATG 786  
 269 LysIleArgGlyAlaHisAsnTyrSerAsnAlaLeuAlaLeuAlaLeuGlyHisAla 288  
 787 AAATCTCCCGGCGACGATAACTACACCAATCGCGTGGCGCGCTGGCGATGATGCT 846  
 289 ValGlyLeuProPheAspAlaMetLeuGlyAlaLeuLysAlaPheSerGlyLeuAlaHis 308  
 847 GCAGGGTTACCGGTGCCGCGCACAGCCTGAAAGCGTTTACCACATTCATGCTGCTGCCGCAT 906  
 309 ArgCysGlnTrpValArgGluArgGlnGlyValSerTyrAspAspSerLysAlaThr 328  
 907 CGCTTTGAAGTTGTGCTGGAGCATACCGCGCTAGCTTGGATTAAACGATTTCGAAAGCGACC 966  
 329 AsnValGlyAlaAlaLeuAlaAlaIleGluGlyLeuGlyAlaAspIleAspGlyLysLeu 348  
 967 AACCTCGGCATACGGAAGCGCGCTGAATGGCGTG-----CACGTAGACGCGACACTG 1020  
 349 ValLeuLeuAlaGlyGlyAspGlyLysGlyAlaAspPheHisAspLeuArgGluProVal 368  
 1021 CATTTGTTGCTGGTGGCGATGTTAAATCGCGCGACTTT-----AGCCCACTG 1068  
 369 AlaArgPheCysArgAla-----ValValLeuLeuGlyArgAspAlaGlyLeu 384  
 1069 GCGCGTTACCTGAATGCCGATAACGTACGCTCTGATTGTTTCGTCGTGACGCGCGCAC 1128  
 385 IleAlaGlnAlaLeuGlyAsnAlaValProLeuValArgValAlaThrLeuAspGluAla 404  
 1129 CTGCGC---GCGCTACGCCCGGGAAGTG---GCAGAACAAACCGAAACTATGGAACAGCGC 1182  
 405 ValArgGlnAlaAlaGluLeuAlaArgGluGlyAspAlaValLeuLeuSerProAlaCys 424  
 1183 ATGCGCTGTGCTGCTCGCGTGTTCACGCCGCGGCGATATGTTCTGCTCTCCACAGCGCTG 1242  
 425 AlaSerLeuAspMetPheLysAsnPheGluGluArgGlyArgLeuPheLysAlaVal 444  
 1243 GCCAGCCTTATCAGTTCAAGAACTTTGAACACAGGAGCAATGAGTTTGGCCCTCTCGCGC 1302  
 445 GluGluLeu 447  
 1303 AAGGAGTTA 1311

RESULT 6  
 AAF28548/c  
 ID AAF28548 standard; DNA; 96109 BP.  
 XX  
 AC AAF28548;  
 XX  
 DT 04-APR-2001 (first entry)  
 XX  
 DE Genomic fragment #35.

XX	Genomic library; bacteria; human upper airway; otitis media; sinusitis;
KW	bronchopulmonary; endocarditis; meningitis; ss.
XX	
OS	Moraxella catarrhalis.
XX	
PN	WO200078968-A2.
XX	
PD	28-DEC-2000.
XX	
PF	16-JUN-2000; 2000WO-USI6649.
XX	
PR	18-JUN-1999; 99US-0140121.
XX	
PA	(INCY-) INCYTE GENOMICS INC.
XX	
PI	Lagace RE, Patterson C, Berg KL;
XX	
DR	WPI; 2001-041427/05.
XX	
PT	Genomic library for identifying diagnostic and therapeutic
PT	compositions, and for identifying virulence factors, regulatory
PT	elements and drug targets, comprises Moraxella catarrhalis nucleic
PT	acids -
XX	
PS	Claim 1; Page 345-368; 545pp; English.
XX	
CC	The present invention relates to a Moraxella catarrhalis genomic library
CC	comprising of a combination of 41 nucleic acid molecules (see
CC	AAF28514-AAF28554). The library has a number of uses described in the
CC	specification e.g. is useful for identifying diagnostic and therapeutic
CC	compositions. M. catarrhalis (Branhamella catarrhalis) is a large
CC	aerobic, gram-negative diplococcus, normally found among the bacterial
CC	flora of human upper airways. M. catarrhalis is known to cause acute,
CC	localised infections such as otitis media, sinusitis and bronchopulmonary
CC	infection and life-threatening, systemic diseases including endocarditis
CC	and meningitis.
XX	
SQ	Sequence 96109 BP; 28783 A; 18910 C; 20341 G; 28075 T; 0 other;
Alignment Scores:	
Pred. No.:	1.17e-72 Length: 96109
Score:	889.50 Matches: 213
Percent Similarity:	60.43% Conservative: 68
Best Local Similarity:	45.81% Mismatches: 147
Query Match:	39.36% Indels: 37
DB:	22 Gaps: 12
US-09-701-229-2 (1-448) x AAF28548 (1-96109)	
Qy	12 ValValGlyLeuGlyLysSerGlyMetSerLeuValargTyrLeuAlaAargArgGlyLeu 31
Db	12721 GTCGTCGGTTGGGCACCTCTGGACCTGTCAGCTGTGNAATTTTGGTCAATCAGGTCAC 12666
Qy	32 ProPheAlaValValAspThrArgGluAsnProGluLeuAlaThrLeuArgAlaGln 51
Db	12661 AAAGTTAGCATCATCGATGGTAAATCCCAATCCT-----ACCTTGGCAGATAAA 12614
Qy	52 TyrProGln---ValGluValArgCysGlyGluLeuAspAlaGluPheLeuCysSerAla 70
Db	12613 CTGCCCCACAGCGGCTTCAATTCATTTGGCAGCTTGGACAGTCTTAAAGTCAAGTCAAGC 12554
Qy	71 ArgGluLeuTyrValSerProGlyLeuSerLeuArgThrProAlaLeuValGlnAlaLa 90
Db	12553 GATCAGATCATCATCACCCCGGCATTTAACCCCAAGAACGCCGAGTAFTTGTGCGCGCCAAA 12494
Qy	91 AlalysGlyValArgIleSerGlyAspIleAspLeuPheAlaAargGluAlalysAla--- 109
Db	12493 GCTCAAGGCATTCTCTGTGATCAGCATGTTCAATTTGTTTCATTGATACACTCAAGACAGA 12434
Qy	110 -----ProIleValAlaIleThrGlySerAsnAlalysSer 121
Db	12433 GATCTAACTCAAGCAACAAACACACCAATCATCGTCATCAGGCTCAATGCGCAAGGC 12374



```
QY 12 ValValGlyLeuGlyLysSerGlyMetSerLeuValArgTyrLeuAlaArgArgGlyLeu 31
   :::::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db 31 ATCATCGGGCTGGCAAAACAGGTCTTCTGTGTGGATTATCTCTTATCCCAACAGGCT 90
   :::::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
QY 32 PropheAlaValValAspThrArgGluAsnProProGluLeuAlaThrLeuArgAlaGln 51
   :::::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db 91 AATATTGCTGTGATTGATATCGAAATAATCTCTGTTGTTATGATAAATCTT----- 141
   :::::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
QY 52 TyrProGln---ValGluValArgCysGlyGluLeuAspAlaGluPheLeuCysSerAla 70
   :::::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db 142 ---CCTCAAAATATCCCTCTTCATCTGAGTGTGTTAAATCAGGAATGTTACTTGAAAGC 198
   :::::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
QY 71 ArgGluLeuTyrValSerProGlyLeuSerLeuArgThrProAlaLeuValGlnAlaAla 90
   :::::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db 199 GATATGATTCTTATTAGCCAGGCTTGCAGTAAACACACAGAAATTCAAACCGCACTT 258
   :::::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
QY 91 AlalysGlyValArgIleSerGlyAspIleAspLeuPheAlaArgGluAlaLysAlaPro 110
   :::::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db 259 AAAGCGGGAGTGGAAAGTAAATCGCGATATTGAATATTCTCGCGCGCAGCAGCAAGCCA 318
   :::::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
QY 111 IleValAlaIleThrGlySerAsnAlaLysSerThrValThrThrLeuValGlyGluMet 130
   :::::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db 319 ATTGTGGGATTACAGGTTCAATGTTAAAGTAAACCGTAACCTACTTGTAGTTTATGAATG 378
   :::::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
QY 131 AlaValAlaAlaAspLysArgValAlaValAlaGlyGlyAsnLeuGlyThrProAlaLeuAsp 150
   :::::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db 379 GCGAAAGCTGCTGGTGTGAAGTGTGTATGCGCGGAAATATTGGGATTCCTCGCTTGTCA 438
   :::::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
QY 151 LeuLeuAlaAspAspIleGluLeuTyrValLeuGluLeuSerSerPheGlnLeuGluThr 170
   :::::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db 439 TTGTTGAATGAAGATTGTGAACCTTATGTACTAGAGCTTTCTAGTTTTTCAGCTTGACAC 498
   :::::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
QY 171 CysAspArgLeuAsnAlaGluValAlaThrValLeuAsnValSerGluAspHisMetAsp 190
   :::::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db 499 ACTTATAGCTTAAAGCTGCGGAGCGACTGTCTTGAACGTGACTGAAGATCATATGGAT 558
   :::::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
QY 191 ArgTyrAspGlyMetAlaAspTyrHisLeuAlalysHisArgIlePheArgGlyAlaArg 210
   :::::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db 559 CGCTATATGATTAGAGATTATCGCAAGCAAAATATACGCATTTATCATATATGCTAAA 618
   :::::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
QY 211 GlnValValValAsnArgAlaAspAlaLeuThr-----ArgProLeuIleAla 226
   :::::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db 619 GTAGGTGTGTGAACATGAAGTACTGCTGACTTTTGGGAAACGAAATCAAGCGAAA 678
   :::::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
QY 227 AspThrValProCysIrpSerPheGlyLeuAsnLysProAspPheLysAlaPheGlyLeu 246
   :::::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db 679 CATACCTT-----TCTTTTGGGAAATAGTGGGAT-----TATTGGCTA 720
   :::::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
QY 247 IleGluGluAspGlyGlnLysTrpLeuAlaPheGlnPheAspLysLeuLeuProValGly 266
   :::::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db 721 AAAACTGAAATGCAAGCAATATTAAATGTTAAAGATGAAGTATTTACCTTGTGAA 780
   :::::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
QY 267 GluLeuLysIleArgGlyAlaHisAsnTyrSerAsnAlaLeuAlaAlaLeuAlaLeuGly 286
   :::::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db 781 GAAGCTACATTTGTTGGTCCCAATAATATATATGACATTTTGGCAGCAACAGCTGGCA 840
   :::::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
QY 287 HisAlaValGlyLeuProPheAspAlaMetLeuGlyAlaLeuLysAlaPheSerGlyLeu 306
   :::::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db 841 CAGCTATAGTATTAAATTAGATTCAATTCCTGACCGCACTTCTGTCATTTCAAGGGTTA 900
   :::::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
QY 307 AlaHisArgCysGlnTrpValArgGluArgGlnGlyValSerTyrAspAspSerLys 326
   :::::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db 901 GATCATGTTTCAATTAGTGCATCAAGCTAATGGCATTCTGTTGGATTATGACTCTAAA 960
   :::::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
QY 327 AlaThrAsnValGlyAlaAlaLeuAlaAlaIleGluGlyLeuGlyAlaAspIleAspGly 346
   :::::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db 961 GCAACAAATGCGGGAGTACAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1014
   :::::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
QY 347 LysLeuValLeuLeuAlaGlyGlyAspGlyLysGlyAlaAspPheHisAspLeuArgGlu 366
   :::::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db 1015 AAATTGCTATTGTTGTAGCGGAGCAGGAAAGGGCTGATTTTTCAGAAATTAGTGAA 1074
   :::::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
```

**Alignment Scores:**

Pred. No.:	5.64e-70	Length:	1830121.
Score:	879.00	Matches:	206
Percent Similarity:	60.72%	Conservative:	63
Best Local Similarity:	46.50%	Mismatches:	136
Query Match:	38.89%	Indels:	38
DB:	17	Gaps:	10

US-09-701-229-2 (1-448) x AAT42063 (1-1830121)

Qy	12	ValValcGlyLeuGlyLysSerGlyMetSerLeuValArgTyrLeuAlaThrLeuArgAlaGln	31
Db	1203162	ATCATCGGGCTTGGCAAAACAGGCTCTTCTGTGTGATTTATCTCTATCCCAACAGGCT	1203221
Qy	32	ProPheAlaValValAspThrArgGluAAsnProGluLeuAlaThrLeuArgAlaGln	51
Db	1203222	AATATTCGTGTGATGTGATATCTCGAAAAATCCCTACTGGTATTGATAAAGCTT	1203272
Qy	52	TyrProGln---ValGluValArgCysGlyGluLeuAspAlaGluPheLeuCysSerAla	70
Db	1203273	---CCTCAAAATATCCCTCTCATCATCGGTAGTTAAATCATCAGGAAGGTACTTAC	1203329
Qy	71	ArgGluLeuTyrValSerProGlyLeuSerLeuArgThrProAlaLeuValGlnAlaLa	90
Db	1203330	GATATGATCTGTTATAGCCACAGGCTTGGCGTAAAAACACACAGAAATTTCAAAC	1203389
Qy	91	AlaLysGlyValArgIleSerGlyAspIleAspLeuPheAlaArgGluAlaLysAlaPro	110
Db	1203390	AAACGGGAGTGGAGTATCGCGGATTTGAATTTATTCGCCGCCGACGCAAGGCCA	1203449
Qy	111	IleValAlaIleThrGlySerAsnAlaLysSerThrValThrThrLeuValGlyGluMet	130
Db	1203450	ATTGTGGGATTTACAGGTTCAATGGTTAAAGTACCGTAACACTACTTATTGATGAATG	1203509
Qy	131	AlaValAlaAlaAspLysArgValAlaValAlaGlyLysLeuLeuGlyThrProAlaLeuAsp	150
Db	1203510	GCAAAAGCTCTCGTGTGAAGATTTGTTATGGCGGAAATATTGGGATTCGCCGTGTCA	1203569
Qy	151	LeuLeuAlaAspAspIleGluLeuTyrValLeuGluLeuSerSerPheGlnLeuGluThr	170
Db	1203570	TTGTTGTAATGAAGATTGTGAACCTTTATGTACTAGAGCTTTCAGTTTTCAGCTTGAGACA	1203629
Qy	171	CysAspArgLeuAsnAlaGluValAlaThrValLeuAsnValSerGluAspHisMetAsp	190
Db	1203630	ACTTATAGCTTAAAGCTGCGCAGCAGCTGTCTTGAACGTGACTGAAGATCATATGGAT	1203689
Qy	191	ArgTyrAspGlyMetAlaAspTyrHisLeuAlaLysHisArgIlePheArgGlyAlaArg	210
Db	1203690	CGCTATATGATTTAGAACATTTACGCCAAGCAAAATTTACGCTATTTATCATATGCTAA	1203749
Qy	211	GlnValValValAsnArgAlaAspAlaLeuThr-----ArgProLeuIleAla	226
Db	1203750	GTAGGTGTGTTGAACATGAGATAGCTGTCTTTTGGGGGAAACGAAATCAGCGAAA	1203809
Qy	227	AspThrValProCysTrpSerPheGlyLeuAsnLysProAspPheLysAlaPheGlyLeu	246
Db	1203810	CATACCGTT-----TCTTTTGGGAAAATAGTCCGAT-----TATTGGCTTA	1203851
Qy	247	IleGluAluAspGlyGlnLysTrpLeuAlaPheGlnPheAspLysLeuLeuProValGly	266
Db	1203852	AAAACCTGAAATGGCAAGCAATATTTAATGGTAAAAAGATGAAGTGATTTTACCTTGTGA	1203911
Qy	267	GluLeuLysIleArgGlyAlaHisAsnTyrSerAsnAlaLeuAlaLeuAlaLeuGly	286
Db	1203912	GAAGCTACATTTGTTGGTCCCAATATTTATGAACATTTTGGCAGCACACATTTGGCA	1203971
Qy	287	HisAlaValGlyLeuProPheAspAlaMetLeuGlyAlaLeuLysAlaPheSerGlyLeu	306
Db	1203972	CAAGCTATAGGTATTAATTAGATTCAATTCGTACCGCTTCGTCATTTCAAAGGGTTA	1204031
Qy	307	AlaHisArgCysGlnTrpValArgGluArgGlnGlyValSerTyrTrpAspAspSerLys	326
Db	1204032	GATCATCGTTTCAATTAGTGTATCAAGCTAATGATGATTCGTTGGATTAATGACTTAA	1204091

Qy	327	AlaThrAsnValGlyAlaAlaLeuAlaAlaLeuGluGlyLeuGlyAlaAspIleAspGly	346
Db	1204092	GCACAAATGTGGGAGTACATGTGCTGCTGGGCTT-----TATATTGAGGCT	1204135
Qy	347	LysLeuValLeuAlaGlyGlyAspGlyLysGlyAlaAspPheHisAspLeuArgGlu	366
Db	1204146	AAATGTCATTGTGTAGCGGAGACGGAAAGGGCTGATTTTCAGAAATTAGCTGAA	1204205
Qy	367	ProValAlaArg-----PheCysArgAlaValValLeuLeuGlyArgAsp	381
Db	1204206	TTAATTAATCAACCACACATTATTGTTATTGT-----TTTGTCGAGAT	1204250
Qy	382	AlaGlyLeuIleAlaGlnAlaLeuGlyAsnAlaValProLeuValArgValAlaThrLeu	401
Db	1204251	GGTGGCGTGTCTGCAAAATTTTCATCGCAAAAGTTAT-----TTG	1204289
Qy	402	AspGluAlaValArgGlnAlaAlaGluLeuAlaArg-----GluGlyAspAla	417
Db	1204290	TTCGATACATGGAAACAGCGATAGAAATTTTACGCCCAACATTCGAAACCGGAGATATG	1204349
Qy	418	ValLeuLeuSerProLacysAlaSerLeuAspMetPheLysAsnPheGluGluArgGly	437
Db	1204350	GTATTATTGTCGCTGCTGTGCAAGTCTCGATCAGTTTCTTCTTTTGAAGAAGCGCGC	1204409
Qy	438	ArgLeuPhe 440	
Db	1204410	GAAGATTT 1204418	
RESULT	9		
IDA	ABA92787/C		
ID	ABA92787 standard; DNA; 640681 BP.		
XX	ABA92787;		
XX	27-MAR-2002 (first entry)		
XX	Buchnera sp. genomic DNA SEQ ID NO:1.		
DE	Buchnera; cockroach-symbiotic bacterium; cockroach extermination;		
XX	circular; ds.		
KW	Buchnera sp.		
OS	Buchnera sp.		
XX	JF2001292771-A.		
XX	23-OCT-2001.		
XX	07-APR-2000; 2000JP-0107160.		
XX	07-APR-2000; 2000JP-0107160.		
XX	(RIKA ) RIKAGAKU KENKYUSHO.		
XX	WPI; 2002-126043/17.		
XX	A genomic DNA of cockroach-symbiotic bacterium -		
PS	Claim 1; Page 16-230; 237pp; Japanese.		

XX  
DE Buchnera sp. genomic DNA SEQ ID NO:1.

CC resultant culture. The DNA is useful for developing agricultural  
 CC chemicals for exterminating cockroaches. The present sequence represents  
 CC the specifically claimed Buchnera sp. genomic DNA sequence, from the  
 CC present invention.

XX  
 SQ Sequence 640681 BP; 237522 A; 83822 C; 84757 G; 234580 T; 0 other;

# Alignment Scores:

Pred. No.: 4.74e-55 Length: 640681  
 Score: 716.00 Matches: 151  
 Percent Similarity: 57.50% Conservative: 102  
 Best Local Similarity: 34.32% Mismatches: 173  
 Query Match: 31.68% Indels: 14  
 DB: 24 Gaps: 7

US-09-701-229-2 (1-448) x ABA92787 (1-640681)

Qy 11 IleValValGlyLeuGlyLysSerGlyMetSerLeuValArgTyrLeuAlaArgArgGly 30  
 Db 241263 TTAATTTAGGAATGGGTTTAAACAGGAATATCTTGCATTATATTTTAAAAAAGGA 241204  
 Qy 31 LeuProPheAlaValAlaAspThrArgGluAsnProGluLeuAlaThrLeuArgAla 50  
 Db 241203 ATAAACCTAAATATATGATGATCTAAACATCCTTCARATTTTATTAATTT 241150  
 Qy 51 GlnTyrProGln---ValGluValArgCysGlyGluLeuAspAlaGluPheLeuCysSer 69  
 Db 241149 -----CCTCAAAATATGATATAGTTAGGAGTTTACACCATCAGTGATCTAGAA 241096  
 Qy 70 AlaArgGluLeuTyrValSerProGlyLeuSerLeuArgThrProAlaLeuValGlnAla 89  
 Db 241095 TCAGACTTAATGTTATAGTCCGGTATTTTCATTTAAACCTATTTTAAATTAAGCA 241036  
 Qy 90 AlaAlaLysGlyValArgIleSerGlyAspIleAspLeuPheAlaArgGluAlaLysAla 109  
 Db 241035 CGTTTGTAGTATGATATATAGTATGATATGAACTATTTCTAGAGAGTACCTGT 240976  
 Qy 110 ProIleValAlaIleThrGlySerAsnAlaLysSerThrValThrLeuValGlyGlu 129  
 Db 240975 CCTATTATTTCAATAACAGGAGCTAATGGAAGAGTACTGTAGCTACCATGATGAAAA 240916  
 Qy 130 MetAlaValAlaAlaAspLysArgValAlaValGlyGlyAsnLeuGlyThrProAlaLeu 149  
 Db 240915 ATTGCAAAAATCAGGATATAAGCTTTTCTAGTGGTATATAGTATGTTCCCGTACTA 240856  
 Qy 150 AspLeuLeuAlaAspAspIleGluLeuTyrValLeuGluLeuSerSerPheGlnLeuGlu 169  
 Db 240855 GAAATCTTGATGAAGAGCAGATTTATACATATAGCACTGTCTAGTTTCAACTAGAA 240796  
 Qy 170 ThrCysAspArgLeuAsnAlaGluValAlaThrValLeuAsnValSerGluAspHisMet 189  
 Db 240795 AATACATTTAATTTAAATCTAAGATAGCAGTTATTTCTAATATAGCAAGATCATATC 240736  
 Qy 190 AspArgTyr---AspGlyMetAlaAspTyrHisLeuAlaLysHisArgIlePheArgGly 208  
 Db 240735 ATCGATACCAATGATGATCCCAACAATAATAAACAATAATGCTGTGTTTAAATCAAA 240676  
 Qy 209 AlaArgGlnValValAlaAsnArgAlaAspAlaLeuThrArgProLeuIleAlaAspThr 228  
 Db 240675 GCAGAAATTTCTATATTAATCTAATGATAAGATACAAAAAGTCTTATTCATAGTAAG 240616  
 Qy 229 ValProCysTrp---SerPheGlyLeuAsnLysProAspPheLysAlaPheGlyLeuIle 247  
 Db 240615 AATAAAAAATGGATTAGTTTGGCAACAATAAGAGTATTCGT-----ATTGT 240565  
 Qy 248 GluCluAspGlyGlnLysTrpLeuAlaPheGlnPheAspLysLeuLeuProValGlyGlu 267  
 Db 240564 TCTAAAGCAATGACCTATTTTATTTTAAAAATAAAAAATATTAATACTAGTACAA 240505  
 Qy 268 LeuLysIleArgGlyAlaHisAsnTyrSerAsnAlaLeuAlaLeuLeuGlyHis 287  
 Db 240504 ATATTATATATGATATCATATTAATTAATAATATATATATATATATATATATATATAT 240445

Qy 288 AlaValGlyLeuProPheAspAlaMetLeuGlyAlaLeuLysAlaPheSerGlyLeuAla 307  
 Db 240444 GCAATGCAATCCCTAGAAATGACCAATAAATGACTTAAAGTTTCTCGAATTTACCA 240385  
 Qy 308 HisArgCysGlnTyrPValArgGluArgGlnGlyValSerTyrTyrAspSerLysAla 327  
 Db 240384 CATCGATTTTCAATAATAAATAAAGAGGAGTACGTTGGATAAAGGATCTAAATCC 240325  
 Qy 328 ThrAsnValGlyAlaAlaLeuAlaLysLeuGlyLeuGlyAlaAspIleAspGlyLys 347  
 Db 240324 ACAATGTAATAGTACTCAAGTTGCTTTAAATCTCTATA-----AAACAACGGAACA 240271  
 Qy 348 LeuValLeuLeuAlaGlyGlyAspGlyLysGlyAlaAspPheHisAspLeuArgGluPro 367  
 Db 240270 ATAAGATTATTTAGCGCGGATAGTAAATCTGCAATTTTAAATATATATAAATAATTT 240211  
 Qy 368 ValAlaArgPheCysArgAlaValValLeuLeuGlyArgAspAlaGlyLeuIleAlaGln 387  
 Db 240210 TTTAGAACTTTTAAAAATAAATAATTTACTGTTTGGAGAGATGGCATTTAAATTTGCAAAA 240151  
 Qy 388 AlaLeuGlyAsnAlaValProLeuValArgValAlaThrLeuAspGluAlaValArgGln 407  
 Db 240150 ATATGTGAAAAA-----AATCTATTTATGTTGAAAAATTTAAAAAAGCAGTATTTTA 240097  
 Qy 408 AlaAlaGluLeuAlaArgGluGlyAspAlaValLeuLeuSerProAlaCysAlaSerLeu 427  
 Db 240096 ATTTCAAAACAAGTCAATCAGTATGATACCTGTTCTTTGCTCTGCTGGTGCAGCAGCTTA 240037  
 Qy 428 AspMetPheLysAsnPheGluGluArgGlyArgLeuPheAlaLysAlaValGluLeu 447  
 Db 240036 GGTCATTTTCTAATCTTTGAAGAGAGAGGTTAATCTTTTATATAAATAATTAAGGAATA 239977  
 RESULT 10  
 AAZ53175/C  
 ID AAZ53175 standard; DNA; 1380 BP.  
 XX AAZ53175;  
 AC AAZ53175;  
 DT 21-MAR-2000 (first entry)  
 DE Neisseria meningitidis ORF 090 partial DNA sequence SEQ ID NO:301.  
 XX Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;  
 KW antigenic; diagnosis; immunogenic; infection; meningitis; septicaemia;  
 KW antibacterial; gene therapy; ds.  
 OS Neisseria meningitidis.  
 XX  
 FN W09957280-A2.  
 XX  
 PD 11-NOV-1999.  
 XX  
 PF 30-APR-1999; 99WO-US09346.  
 XX  
 PR 01-MAY-1998; 98US-0083758.  
 PR 31-JUL-1998; 98US-0094869.  
 PR 02-SEP-1998; 98US-0098994.  
 PR 02-SEP-1998; 98US-0099062.  
 PR 09-OCT-1998; 98US-0103749.  
 PR 09-OCT-1998; 98US-0103794.  
 PR 09-OCT-1998; 98US-0103796.  
 PR 25-FEB-1999; 99US-0121528.  
 XX  
 PA (CHIR ) CHIRON CORP.  
 PA (GENO-) INST GENOMIC RES.  
 XX  
 PI Fraser C, Galeotti C, Grandi G, Hickey E, Massignani V, Mora M;  
 PI Petersen J, Pizsa M, Rappuoli R, Ratti G, Scalato E, Scarselli M;  
 PI Tettelin H, Venter JC;  
 XX  
 DR WPI: 2000-062150/05.  
 DR P-PSDB: AAY74413.  
 XX





PR 30-APR-1999; 99US-0132068.  
 XX (CHIR ) CHIRON CORP.  
 XX Frazer CM, Hickey E, Peterson J, Tettelin H, Venter JC;  
 PI Masignani V, Galeotti C, Mora M, Ratti G, Scarselli M, Scariato V;  
 PI Rappuoli R, Pizza M;  
 XX WPI; 2000-318079/27.  
 XX Isolated nucleotide sequences of *Neisseria meningitidis* which can be  
 PT used in the diagnosis and treatment of *N. meningitidis* infection and  
 PT other *Neisseria* infections, for example, *N. gonorrhoea* -  
 XX Claim 7; Page 547-567; 1760pp; English.  
 XX The present invention describes methods of obtaining immunogenic  
 CC proteins from *Neisseria* genomic sequences. AA81453 to AA82414  
 CC represent specifically claimed *Neisseria meningitidis* genomic DNA  
 CC sequences; AA81260 to AA81303 and AA825620 to AA825663 represent  
 CC *Neisseria* DNA sequences and their corresponding proteins; AA81254 to  
 CC AA81259 and AA81304 to AA81321 represent PCR primers used in the  
 CC isolation of *Neisseria meningitidis* DNA sequences; and AA81322 to  
 CC AA81452 represent *Neisseria meningitidis* MenB polynucleotide ORF  
 CC sequences, which are all used in the exemplification of the present  
 CC invention. The nucleic acid sequences, protein sequences, and antibodies  
 CC against them, can be used in the manufacture of a composition. The  
 CC composition can be used as a medicament (or in the manufacture of a  
 CC medicament) for treating, preventing or diagnosing infection due to  
 CC *Neisseria* bacteria. For example, some of the identified proteins could  
 CC be components of vaccines against *Meningococcus B*; against all serotypes;  
 CC and/or against all pathogenic *Neisseriae*. Identification of sequences  
 CC from the bacterium will also facilitate production of biological probes,  
 CC particularly organism-specific probes. Attempts to make efficacious  
 CC *Meningococcus B* vaccines have failed mainly due to antigen tolerance.  
 CC Multivalent vaccines have also been tried but none have successfully  
 CC overcome antigenic variability. The provision of further, complete  
 CC sequences may provide an opportunity to identify secreted or surface  
 CC exposed proteins that may be presumed targets for the immune system and  
 CC which are not antigenically variable or at least more conserved than  
 CC other more variable regions.  
 XX  
 SQ Sequence 69936 BP; 16512 A; 19259 C; 17399 G; 16763 T; 3 other;  
 Alignment Scores:  
 Pred. No.: 3,58e-55 Length: 69936  
 Score: 704.00 Matches: 173  
 Percent Similarity: 53.52% Conservatives: 70  
 Best Local Similarity: 38.11% Mismatches: 177  
 Query Watch: 31.15% Indels: 34  
 DB: 21 Gaps: 8  
 US-09-701-229-2 (1-448) x AA81479 (1-69936)  
 QY 11 IleValValGlyLeuGlyLysSerGlyMetSerLeuValArgTyrLeuAlaArgGly 30  
 DB 52618 CTCGTCGCCGACTCGCGGTACGGGTATTTCCATGATTCCTCGCGCAAAACGGC 52559  
 QY 31 LeuProPheAlaValAlaValAspThrArgGluAsnProProGluLeuAlaThrLeuArgAla 50  
 DB 52558 CGCGAGGTGCTCGGTATGATGGGAGCTGAGCCGGAACCGCGTGTCCGAATCGGTAAAG 52499  
 QY 51 GlnTyrProGlnValGluAlaValArgCysGlyGluLeuAspAlaGluPheLeuCysSerAla 70  
 DB 52498 ATGTTTACGGGTGTGGTGTATTTACACGGCGCGTGAAGATCGCTGGACACGGTTTC 52439  
 QY 71 ArgGluTyrValSerProGlyLeuSerLeuArgThrProAlaLeuGlnAlaAla 90  
 DB 52438 GATATTCGTGCTCTCAGTCCCGGCATCAGCGAGCGCGCGGATATTGAGCGCTTCAAG 52379  
 QY 91 AlaLysGlyValArgIleSerGlyAspIleAspLeuPheAla-----Arg 105  
 DB 52378 CAAAACGGCGAGCGGTGTTGGGCGACATCGAATTGCTGCGGACATTTGTGAACCGCGCG 52319

QY 106 GluAlaLysAlaProIleValAlaIleThrGlySerAsnAlaLysSerThrValThrThr 125  
 DB 52318 GAGCAGCAAG-----GTAATTGGATACCGCGCAGCAACGCAACCGGTAAACGAGC 52265  
 QY 126 LeuValGlyGluMetAlaValAlaAlaAspLysArgValAlaValGlyGlyAsnLeuGly 145  
 DB 52264 CTGGTCGGCTATCTCTGTATCAAGTGGCGGTGATACCTGATCGCGGCAATATCGGC 52205  
 QY 146 ThrProAlaLeuAspLeu-----LeuAlaAspAspIleGluLeuValLeu 161  
 DB 52204 ACSCGGGTTTGGAGCGGGAATGCGACGGGAGGCAAAAGCGGAGGTGGTGGTGTG 52145  
 QY 162 GluLeuSerSerPheGlnLeuGluThrCysAspArgLeuAsnAlaGluValAlaThrVal 181  
 DB 52144 GAGCTTCCAGCTTCCAACTGGAAACACCGCAAGCGCTGCGTCCGAGTGGCGGCGAGT 52085  
 QY 182 LeuAsnValSerGluAspHisMetAspArgTyrAspGlyMetAlaAspTyrHisLeuAla 201  
 DB 52084 CTGAACATTTCCGAAGACCATCTCGACCGCTACGACGACTGTCTCGACTATGCGCATACC 52025  
 QY 202 LysHisArgIlePheArgGlyAlaArgGlnValValValAsnArgAlaAspAlaLeuThr 221  
 DB 52024 AAAGCCAGATTTTCCGTGCGCGCGGTGAGTGTGAGTGGCGGAGTGTGCTTC--- 51968  
 QY 222 ArgProLeuIleAlaAspThrValProCysTrpSerPheGlyLeuAsnLysProAspPhe 241  
 DB 51967 -----TGCGCGCGATGAAGCGTGCCTGCGCGCGGAGGTA 51935  
 QY 242 LysAlaPheGlyLeuIleGluAspGlyGlnLysTrpLeuAlaPheGlnPhe----- 259  
 DB 51934 AAATGGTGTTCGTTG---GAACAGCAAGCTGATTTCTGGTGGACCGAGGAGCGCGC 51878  
 QY 260 -----AspLysLeuLeuProValGlyGluLeuLysIleArgGlyAlaHis 274  
 DB 51877 CTGAACACAGCAATGAAGATTTGTTGTACGCAACACATTCGTTGCAAGGTCTGCAC 51818  
 QY 275 AsnTyrSerAsnAlaLeuAlaLeuLeuGlyHisAlaValGlyLeuProPheAsp 294  
 DB 51817 AACCGCGCTAACGATGCTGCGTGTGTTGTGAGGCGCATCGTTTCTCGCGGAA 51758  
 QY 295 AlaMetLeuGlyAlaLeuLysAlaPheSerGlyLeuAlaHisArgCysGlnTrpValArg 314  
 DB 51757 GCATTTGCTCGAACACGCTCCAAACCTCCAGGCGTGCCTGCGCGGAAAAAATCGGC 51698  
 QY 315 GluArgGlnGlyValSerTyrTyrAspSerLysAlaThrAsnValGlyAlaAlaLeu 334  
 DB 51697 GAGAAAAACGGCGTGTGTTTATCGACGACAGCAAGCGCATGTTCGGCGCGACTGCC 51638  
 QY 335 AlaAlaIleGluGlyLeuGlyAlaAspIleAspGlyLysLeuValLeuAlaGlyGly 354  
 DB 51637 GCCCGATTCGCGGTTTG-----CAAAATCGCTCTCTCGTATTTTGGCGGC 51590  
 QY 355 AspGlyLysGlyAlaAspPheHisAspLeuArgGluProValAlaAlaArgPheCysArgAla 374  
 DB 51589 ATGGGTAAAGCGAGGAGTTCACGCCCTGCGCGCATGCTGTTAGGCAAGCGCAAGGCGC 51530  
 QY 375 ValValLeuLeuGlyArgAspAlaGlyLeuIleAlaGlnAlaLeu---GlyAsnAlaVal 393  
 DB 51529 GTGTTCTGATGTTGTCGATGCGCGCAATTCGCGCGATTTGGAGCGGTGGCGCTTG 51470  
 QY 394 ProLeuValArgValAlaThrLeuAspGluAlaValArgGlnAlaAlaGluLeuAlaArg 413  
 DB 51469 AATATGACGACTCGCCACTTTGGGAGAGCGCTTCAGAGCGCATATGCTCCAAAGCGCA 51410  
 QY 414 GluGlyAspAlaValLeuLeuSerProAlaCysAlaSerLeuAspMetPheLysAsnPhe 433  
 DB 51409 GCAGCGATATTGTTGCTCAGCGCGCTGCGGAGCTTTGATATGTTTCAAGGCTAC 51350  
 QY 434 GluGluArgGlyArgLeuPheAlaLysAlaValGluLeuLeu 447  
 DB 51349 GCGCACCGTTCGGAAGTGTATTATCGAAGCGTTTAAAGGCTTTG 51308

## RESULT 12

AAF21607  
ID AAF21607 standard; DNA: 349980 BP.

AC AAF21607;  
XX

DT 13-MAR-2001 (first entry)  
XX

DE Neisseria meningitidis B nucleotide sequence SEQ ID NO:108.  
XX

KW Neisseria meningitidis; Neisseria gonorrhoeae; immunogenic; vaccine;  
KW diagnosis; antigen; detection; infection; gene therapy; antibacterial;  
KW ds.  
XX

OS Neisseria meningitidis.  
XX

PN WO200066791-A1.  
XX

PD 09-NOV-2000.  
XX

PF 08-MAR-2000; 2000WO-US05928.  
XX

PR 30-APR-1999; 99US-0132068.  
XX

PR 08-OCT-1999; 99WO-US23573.  
XX

PR 28-FEB-2000; 2000GB-0004695.  
XX

XX (CHIR ) CHIRON CORP.  
PA

PA (GENO-) INST GENOMIC RES.  
XX

PI Pizza M, Hickey E, Peterson J, Tettelin H, Venter JC, Maignani V;  
PI Galsotti C, Mora M, Ratti G, Scarselli M, Scariato V, Rappuoli R;  
PI Frazer CM, Grandi G;  
XX

DR WPI; 2000-647603/62.  
XX

PT Neisseria meningitidis B full length genome sequence and open reading  
PT frames are used to detect, treat and prevent Neisserial infections -  
XX

PS Claim 7; Appendix A; 692pp; English.  
XX

CC The present invention describes the full length genome of  
CC Neisseria meningitidis B (NMB). The sequences in AAF21544 and AAF21607  
CC to AAF21613 represent fragments of the NMB genomic sequence, as the  
CC sequence was too long to go in a record on its own it was split into 8  
CC sequences which overlap each other at the beginning and end of each  
CC sequence by 49980 bp (i.e. the last 49980 bp of AAF21544 is repeated at  
CC the beginning of AAF21607, the last 49980 bp of AAF21607 are repeated at  
CC the beginning of AAF21608, and so on). AAF21545 to AAF21588 encode the  
CC Neisseria proteins given in AAF58550 to AAF58593, and AAF21589 to  
CC AAF21606 represent PCR primers which are used in the exemplification of  
CC the present invention. The NMB genome and fragments from it have  
CC antibacterial activity, and can be used in vaccines and gene therapy.  
CC Neisseria nucleic acids, proteins and/or antibodies which binds to the  
CC proteins can be used in compositions for treating or preventing infection  
CC due to Neisserial bacteria or as a diagnostic reagent for detecting the  
CC presence of Neisserial bacteria or of antibodies raised to Neisserial  
CC bacteria. Computers, computer memory, computer storage medium or computer  
CC databases can be used in a search to identify open reading frames (ORFs)  
CC or coding sequences within the NMB genome. The DNA sequences provide  
CC further opportunities to find antigenic or immunogenic proteins which are  
CC more effective in vaccines than the outer membrane proteins currently  
CC used.  
XX

SQ Sequence 349980 BP; 84410 A; 84863 C; 94187 G; 86520 T; 0 other;

## Alignment Scores:

Pred. No.:	2,978-54	Length:	349980
Score:	704.00	Matches:	173
Percent Similarity:	53.52%	Conservative:	70
Best Local Similarity:	38.11%	Mismatches:	177
Query Match:	31.15%	Indels:	34
DB:	21	Gaps:	8

US-09-701-229-2 (1-448) x AAF21607 (1-349980)

QY	11	IleValValcIlyLeuGlyLysSerGlyMetSerLeuValArgTyrLeuAlaAArgGly	30
DB	128569	CTCGTCGCGGACTCGCGGTATTCATGATTCCTACCTCGCAAAACGGC	128628
QY	31	LeuProPheAlaValValAspThrArgGluAsnProProGluLeuAlaThrLeuArgAla	50
DB	128629	CGCGAGGTTGCTGCGTATGATCGGAGCTGAAGCCGAGCGGTGCGCAATCGGTAAG	128688
QY	51	GlnTyrProGlnValGluValArgCysGlyGluLeuAspAlaGluPheLeuCysSerAla	70
DB	128689	ATGTTTTCGCGGTGCGTATTCACGCGCGCTCTGAAGATCGCGTGCACACCGGTTTC	128748
QY	71	ArgGluLeuTyrValSerProGlyLeuSerLeuArgThrProAlaLeuValGlnAlaAla	90
DB	128749	GATATTCTGCTCTCAGTCCCGCATCAGCGGCGGCGGATATTGAGCGGTTCAAG	128808
QY	91	AlaLysGlyValArgIleSerGlyAspIleAspLeuPheAla-----Arg	105
DB	128809	CAAAACGGCGGAGCGGTGTCGCGACATCGAATTCTGCGGACATTCGTAACCCCGG	128868
QY	106	GluAlaLysAlaProIleValAlaIleThrGlySerAsnAlaLysSerThrValThr	125
DB	128869	GACGACAAG-----GTAATTCGATTACCGCAGCAACGCAACACCGTAAAGGAGC	128922
QY	126	LeuValGlyGluMetAlaValAlaAspLysArgValAlaValGlyGlyAsnLeuGly	145
DB	128923	CTGGTCGCGTATCTGTATCAAGTCGCGGTGGATACCGTTATCGCGGCAATATCGGC	128982
QY	146	ThrProAlaLeuAspLeu-----LeuAlaAspIleGluLeuTyrValLeu	161
DB	128983	ACGCGGTTTTCGAGCGGAGTGGCAGCGCAAGGCAAGGCAAGGCGGAGCGGTGTTG	129042
QY	162	GluLeuSerPheGlnLeuGluThrCysAspArgLeuAsnAlaGluValAlaThrVal	181
DB	129043	GAGCTTTCAGCTTCCAACTGGAAACACCGCAAGAGCTCGCTCCGCTCGGCGGAGCGTG	129102
QY	182	LeuAsnValSerGluAspHisMetAspArgTyrAspGlyMetAlaAspTyrHisLeuAla	201
DB	129103	CTGAACATTTCCGAAGACCATCTCGACCGCTACGAGGACTGTCGACTATGCGCATAC	129162
QY	202	LysHisArgIlePheArgGlyAlaArgGlnValValAlaAsnArgAlaAspAlaLeuThr	221
DB	129163	AAAGCAAGATTTTCGTCGCGGCGGTCGAGGTTTGAATTCGCGGACCATCGCTTC---	129219
QY	222	ArgProLeuIleAlaAspThrValProCysTyrSerPheGlyLeuAsnLysProAspPhe	241
DB	129220	-----TGCCCGCGGATGAAGCGTCCCGCGCGCGAGGTA	129252
QY	242	LysAlaPheGlyLeuGluAspGlyGlnLysTrpLeuAlaPheGlnPhe-----	259
DB	129253	AAATGTTTTCGTTG---GAACACGAAGCTGATTTCTGTTGGAACCGCAGACAGCCGC	129309
QY	260	-----AspLysLeuLeuProValGlyGluLeuLysIleArgGlyAlaHis	274
DB	129310	CTGAACAAAGCAAGATTTGATTCACGCAAGACATTCCTCGTGAAGGCTCTGCAAC	129369
QY	275	AsnTyrSerAsnAlaLeuAlaLeuAlaLeuGlyHisAlaValGlyLeuProPheAsp	294
DB	129370	AACGCGGTACGTCATGCTGCGCTGCTTGTGTGAGGCCATCGGTTTTCGCGCGAA	129429
QY	295	AlaMetLeuGlyAlaLeuLysAlaPheSerGlyLeuAlaHisArgCysGlnTrpValArg	314
DB	129430	GCATTGCTCGAACACGCTCAAAACCTTCCAAAGCCTCCGCGCGCGTGAAGAAATCGGC	129489
QY	315	GluArgGlnGlyValSerTyrAspSerLysAlaThrAsnValGlyAlaAlaLeu	334
DB	129490	GAGAAACCGCGCTGCTGTTTATCGACGACGCAAGGCAAGTGTGCGCGCGACTGCC	129549
QY	335	AlaAlaIleGluGlyLeuGlyAlaAspIleAspGlyLysLeuValLeuAlaGlyGly	354
DB	129550	GCCGCGATTGCGCGTTTG-----CAAAATCCGCTCTTCGTGATTGCGCGGC	129597

QY 355 AspGlyGlyAlaAspPheHisAspLeuArgGluProValAlaArgPheCysArgAla 374  
 DB 129598 ATGGGTAAAGGAGGAGGACTTCAGCCCTCGCGATGCACTGGTGGCAAGCAAAAGGC 129657  
 QY 375 ValValLeuLeuGlyArgAspAlaGlyLeuLeuAlaGlnAlaLeu---GlyAsnAlaVal 393  
 DB 129658 GTGTTCTGATTGGTCTCGATGCGCGCAAAATCCGCGGATTTGACGCGCTGCGGCTTG 129717  
 QY 394 ProLeuValArgValAlaThrLeuAspGluAlaValArgGlnAlaAlaGluLeuAlaArg 413  
 DB 129718 AATATGACCGACTGCCCACTTTGGGAGAGAGCGGTTTCAGCGCATATGCCCAAGCCGAA 129777  
 QY 414 GluGlyAspAlaValLeuLeuSerProAlaCysAlaSerLeuAspMetPheTyrAsnPhe 433  
 DB 129778 GCAGGCGATATTGTTGTCAGCCCGCTCGCGAGCTTTGATATGTTCAAAGGCTAC 129837  
 QY 434 GluGluArgGlyArgLeuPheAlaLysAlaValGluGluLeu 447  
 DB 129838 GGCACCGGTTCCGAAGTGTATTATCGAAGCGTTTAAGGCTTTG 129879

RESULT 13

AAA81490  
 ID AAA81490 standard; DNA; 1437668 BP.

XX  
 AC AAA81490;  
 XX  
 DT 04-DEC-2000 (first entry)  
 XX  
 DE N. meningitidis B full length genome DNA sequence SEQ ID NO:1068.  
 XX  
 KW Neisseria meningitidis; Neisseria gonorrhoeae; genome; immunogenic;  
 KW antigen; vaccine; diagnosis; infection; antibacterial; identification;  
 KW Meningococcus B; MenB; ds.  
 XX  
 OS Neisseria meningitidis.  
 XX  
 PN WO200022430-A2.  
 XX  
 PD 20-APR-2000.  
 XX  
 PF 08-OCT-1999; 99WO-US23573.  
 XX  
 PR 09-OCT-1998; 98US-0103794.  
 PR 30-APR-1999; 99US-0132068.  
 XX  
 PA (CHIR ) CHIRON CORP.  
 XX  
 PI Frazer CM, Hickey E, Peterson J, Tettelin H, Venter JC;  
 PI Maignani V, Galeotti C, Mora M, Ratti G, Scarselli M, Scarlato V;  
 PI Rappuoli R, Pizza M;  
 XX  
 DR WPI; 2000-318079/27.

XX Isolated nucleotide sequences of Neisseria meningitidis which can be  
 PT used in the diagnosis and treatment of N. meningitidis infection and  
 PT other Neisserial infections, for example, N.gonorrhoea -  
 XX  
 PS Claim 7; Page 866-1272; 1760pp; English.  
 XX  
 CC The present invention describes methods of obtaining immunogenic  
 CC proteins from Neisseria genomic sequences. AAA81453 to AAA82414  
 CC represent specifically claimed Neisseria meningitidis genomic DNA  
 CC sequences; AAA81260 to AAA81303 and AAB25620 to AAB25663 represent  
 CC Neisseria DNA sequences and their corresponding proteins; AAA81254 to  
 CC AAA81259 and AAA81304 to AAA81321 represent PCR primers used in the  
 CC isolation of Neisseria meningitidis DNA sequences; and AAA81322 to  
 CC AAA81452 represent Neisseria meningitidis MenB polynucleotide ORF  
 CC sequences, which are all used in the exemplification of the present  
 CC invention. The nucleic acid sequences, protein sequences, and antibodies  
 CC against them, can be used in the manufacture of a composition. The  
 CC composition can be used as a medicament (or in the manufacture of a  
 CC medicament) for treating, preventing or diagnosing infection due to

CC Neisserial bacteria. For example, some of the identified proteins could  
 CC be components of vaccines against Meningococcus B; against all serotypes;  
 CC and/or against all pathogenic Neisseriae. Identification of sequences  
 CC from the bacterium will also facilitate production of biological probes,  
 CC particularly organism-specific probes. Attempts to make efficacious  
 CC Meningococcus B vaccines have failed mainly due to antigen tolerance.  
 CC Multivalent vaccines have also been tried but none have successfully  
 CC overcome antigenic variability. The provision of further, complete  
 CC sequences may provide an opportunity to identify secreted or surface  
 CC exposed proteins that may be presumed targets for the immune system and  
 CC which are not antigenically variable or at least more conserved than  
 CC other more variable regions.

XX  
 SQ Sequence 1437668 BP; 344338 A; 353206 C; 385074 G; 355045 T; 5 other;

Alignment Scores:  
 Pred. No.: 1-91e-53 Length: 1437668  
 Score: 704.00 Matches: 173  
 Percent Similarity: 53.52% Conservativity: 70  
 Best Local Similarity: 38.11% Mismatches: 177  
 Query Match: 31.15% Indels: 34  
 DB: 21 Gaps: 8

US-09-701-229-2 (1-448) x AAA81490 (1-1437668)

QY 11 IleValValGlyLeuGlyLysSerGlyMetSerLeuValArgTyrLeuAlaArgGly 30  
 DB 428569 CTGTCGCGGACTCGCGGTATTCATGATTCATGCTACCTACGCAAAAGGC 428628  
 QY 31 LeuProPheAlaValValAspThrArgGluAsnProGluLeuAlaThrLeuArgAla 50  
 DB 428629 GCGAGGTTGCTCGTATGATCGGAGCTGAAGCCGAGCGGTGTCGAAATCGGTAA 428688  
 QY 51 GlnTyrProGlnValGluValArgCysGlyGluLeuAspAlaGluPheLeuCysSerAla 70  
 DB 428689 ATGTTTGACGGTGTGTTTACACGGCGCTGAAAGCTGCGTGGACACGGTTTC 428748  
 QY 71 ArgGluLeuTyrValSerProGlyLeuSerLeuArgThrProAlaLeuGlnAlaAla 90  
 DB 428749 GATATTCTGGCTCTCAGTCCGCGATCAGCGAGCGGATGATGATGATGATGAT 428808  
 QY 91 AlaLysGlyValArgLysSerGlyAspLeuPheAla-----Arg 105  
 DB 428809 CAAACGCGGAGCGGTGTTGGCGGACATCGATGCTGCGGAGACATGTGAACCGCGG 428868  
 QY 106 GluAlaLysAlaProIleValAlaIleThrGlySerAsnAlaLysSerThrValThrThr 125  
 DB 428869 GACGACAAAG-----GTAATGCGATTACCGGACGACGCAAAACCGGTACACG 428922  
 QY 126 LeuValGlyGluMetAlaValAlaAlaAspLysArgValAlaValGlyGlyAsnLeuGly 145  
 DB 428923 CTGTCGCGCTATCTCTATCAAGTGGCGGTGATACCGTATCGCGGGCAATATCGGC 428982  
 QY 146 ThrProAlaLeuAspLeu-----LeuAlaAspLysPheGluLeuTyrValLeu 161  
 DB 428983 ACGCCGTTTTGGAGCGGAAATGCGACGCGGAGGCAAAAGCGCGGTGTTGTTG 429042  
 QY 162 GluLeuSerSerPheGlnLeuGluThrCysAspArgLeuAsnAlaGluValAlaThrVal 181  
 DB 429043 GAGCTTCCAGCTTCCAACTGGAAACACCCGAAAGCTGCGGTGCGGCGGACCGGTG 429102  
 QY 182 LeuAsnValSerGluAspHisMetAspArgTyrAspGlyMetAlaAspTyrHisLeuAla 201  
 DB 429103 CTGAACATTTCCGAGACCATCTCGACCGGTACGACGCTGTCGACTATGCGCATACC 429162  
 QY 202 LysHisArgIlePheArgGlyAlaArgGlnValValValValValValValValValVal 221  
 DB 429163 AAAGCCCAAGATTTTCCGTCGCGACGCGGTGTTGAATGCGGACGATGCGGTTC--- 429219  
 QY 222 ArgProLeuAlaAspThrValProCysTrpSerPheGlyLeuAsnLysProAspPhe 241  
 DB 429220 -----TGCAGCGCGATGAAGCGTGCAGCGCGCGGCGGAGGTA 429252

QY 242 LysAlaPheGlyLeuLeuGluAspGlyGlnLysTrpLeuAlaPheGlnPhe----- 259  
 Db 429253 AAATGTTTTCGTTG---GAACACGAAGCTGATTTCTGGTGGACCGCAGCAGCGCG 429309  
 QY 260 -----AspLysLeuLeuProValGlyGluLeuLysLeuLeuGlyAlaHis 274  
 Db 429310 CTGAACAGCGCAATGAAGATTGTATCCACCAAGACATTCGTTGCAAGGTCTGCAC 429369  
 QY 275 AsnTyrSerAsnAlaLeuAlaLeuAlaLeuGlyHisAlaValGlyLeuLeuProPheAsp 294  
 Db 429370 AAGCGCGCTAAGCTAGCTGCGGTGGCTTGTGTGAGCCATCGTTGTCGCCGGA 429429  
 QY 295 AlaMeLeuGlyAlaLeuLysAlaPheSerGlyLeuAlaHisArgCysGlnTrpValArg 314  
 Db 429430 GCATTCTCGAACACCTCAAAACCTTCCAAAGCGCTCCGACCGCTGGAAAAATCGGC 429489  
 QY 315 GluArgGlnGlyValSerTyrAspAspSerLysAlaThrAsnValGlyAlaLeu 334  
 Db 429490 GAGAAAACGGCGTGTGTATTATCGACGACGACAAAGGACCAATGTCGGCGGACTGCC 429549  
 QY 335 AlaAlaIleGluGlyLeuGlyAlaAspIleAspGlyLysLeuValLeuAlaGlyGly 354  
 Db 429550 GCCGCGATTGCCGTTTG-----CAAAATCCGCTCTCTCGTATTGTCGCGC 429597  
 QY 355 AspGlyLysGlyAlaAspPheHisAspLeuArgGluProValAlaArgPheCysArgAla 374  
 Db 429598 ATGGGTAAAGGCGAGGACTTACCGCCCTGCGGATGCACTGGTAGGCAAGCAAAAGC 429657  
 QY 375 ValValLeuLeuGlyArgAspAlaGlyLeuIleAlaGlnAlaLeu---GlyAsnAlaVal 393  
 Db 429658 GTGTTCTTATGTTGTCATGCGCGCAANTCCCGCGGANTTGGACCGCTGCGGCTTG 429717  
 QY 394 ProLeuValArgValAlaThrLeuAspGluAlaValArgGlnAlaGluLeuAlaArg 413  
 Db 429718 AATATACCGACTGCGCCACTTGGGAGAAGCGTTCACACGCGCATATGCCCAAGCGAA 429777  
 QY 414 GluGlyAspAlaValLeuSerProAlaCysAlaSerLeuAspMetPheLysAsnPhe 433  
 Db 429778 GCAGGCGATATGTTGTCTACGCCCCGCTCGCGAGCTTTGATATGTTCAAAGCTAC 429837  
 QY 434 GluGluArgGlyArgLeuPheAlaLysAlaValGluGluLeu 447  
 Db 429838 GCGCACCGTTCGAAGTGTATTATCGAAGCGTTTAAGGCTTTG 429879  
 RESULT 14  
 ID AB241528 standard; DNA; 1335 BP.  
 AC AB241528;  
 XX  
 XX  
 XX  
 DT 07-MAR-2003 (first entry)  
 DE N. gonorrhoeae nucleotide sequence SEQ ID 7645.  
 KW Antibacterial; infection; vaccine; gene therapy; gene; ds.  
 XX  
 XX  
 OS Neisseria gonorrhoeae.  
 XX  
 PN WO200279243-A2.  
 PD 10-OCT-2002.  
 XX  
 PF 12-FEB-2002; 2002WO-IB02069.  
 XX  
 PR 12-FEB-2001; 2001GB-0003424.  
 XX  
 PA (CHIR-) CHIRON SPA.  
 XX  
 PI Fontana MR, Piza M, Massignani V, Monaci E;  
 XX  
 DR WPI; 2003-058415/05.  
 DR P-PSDB; ABP80558.  
 XX

PT New protein from Neisseria gonorrhoeae, useful for the manufacture of a medicament for treating or preventing N. gonorrhoeae infection -  
 XX  
 PS Disclosure; Page 745-746; 815pp; English.

CC The present invention relates to proteins from Neisseria gonorrhoeae.  
 CC Also disclosed are the nucleic acid molecules encoding the proteins and  
 CC antibodies that specifically bind to the proteins. The composition  
 CC comprising the protein, nucleic acid or antibody is useful for the  
 CC manufacture of a medicament for treating or preventing N. gonorrhoeae  
 CC infection, this may be in the form of a vaccine or gene therapy.  
 CC Sequences given in records AB237706-AB242016 represent nucleic acid  
 CC molecules of the invention.

XX Sequence 1335 BP; 303 A; 341 C; 415 G; 276 T; 0 other;

#### Alignment Scores:

Pred. No.: 3,78e-57 Length: 1335  
 Score: 701.00 Matches: 173  
 Percent Similarity: 52.88% Conservative: 66  
 Best Local Similarity: 38.27% Mismatches: 183  
 Query Match: 31.02% Indels: 30  
 DB: 25 Gaps: 7

US-09-701-229-2 (1-448) x AB241528 (1-1335)

QY 11 IleValValGlyLeuGlyLysSerGlyMetSerLeuValArgTyrLeuAlaArgArgGly 30  
 Db 25 CTCGTGCGCGACTTGGCGCAGGTATTTCCATGATTCCTATCTCGCAAAACGCG 84  
 QY 31 LeuProPheAlaValValAspThrArgGluAsnProProGluLeuAlaThrLeuArgAla 50  
 Db 85 GCGGAGGTTGCCGCTTATGTCGGAGCTCAAAAGCGCGCGTGGCAAAATCGGTAAG 144  
 QY 51 GlnTyrProGluValGluValArgCysGlyGluLeuAspAlaGluPheLeuCysSerAla 70  
 Db 145 ATGTGTGACGGCTGTGTGTTTACAGGCTGCTGCTCAAAAGATCGGTGGCAACAGGTTTC 204  
 QY 71 ArgGluLeuTyrValSerProGlyLeuSerLeuArgThrProAlaLeuValGlnAlaAla 90  
 Db 205 GATATTCTGCGCTCAGTCCCGCATCCAGCGGCGCAGCGCATATCGAGCGGTTCAAG 264  
 QY 91 AlaLysGlyValArgIleSerGlyAspIleAspLeuPheAla-----ArgGluAla 107  
 Db 265 CAAAACGGCGCGCTGTAGCGACATCGAATTGCTGCGGACATTTGTGAACCGCGC 324  
 QY 108 LysAlaProIleValAlaIleThrGlySerAsnAlaLysSerThrValThrThrLeuVal 127  
 Db 325 GCGGACAAGGTGATGCGATTACCGGACGACAGCGCAAAACCCACGTAACGAGCTGGTC 384  
 QY 128 GlyGluMetAlaValAlaAlaAspLysArgValAlaValGlyLysLeuGlyThrPro 147  
 Db 385 GCGTATCTGTCATCAAGTGGCGGTGGATACCGTATTCGTCGCGGCAATATCGTACGCGC 444  
 QY 148 AlaLeuAspLeu-----LeuAlaAspAspIleGluLeuTyrValLeuGluLeu 163  
 Db 445 GTTTTGGAGGAGAAATTCAGCGCAAGCAAAAGCGGACGTGTGGGTGTGGAGCTT 504  
 QY 164 SerSerPheGlnLeuGluThrCysAspArgLeuAsnAlaGluValAlaThrValLeuAsn 183  
 Db 505 TCCAGCTTCCAACTGGAAAACACCGAAGCCCTGCGCCGACTGCGGCGAGGTGATGAAC 564  
 QY 184 ValSerGluAspHisMetAspArgTyrAspGlyMetAlaAspTyrHisLeuAlaLysHis 203  
 Db 565 ATTTCCGAAGACCATCTCGACCGCTACGACGATTTGCTCGACTACGCGCATACCAAGCC 624  
 QY 204 ArgIlePheArgGlyAlaArgGlnValValValAsnArgAlaAspAlaLeuThrArgPro 223  
 Db 625 GAGATTTTCGTCGCGATGAGTGCAGGTTTTGAAT----- 660  
 QY 224 LeuIleAlaAspThrValProCysTrpSerPheGlyLeuAsnLysProAspPheLysAla 243  
 Db 661 -----GCGGACGACGTGTTCTGCCGCCATGAAGCGGCGGCGGTGAAGTGAACGCG 714

Qy	244	PheGlyLeuIleGluGluAspGlyGlnLysTrpLeuAlaPheGlnPhe-----	259
Db	715	TTTTTCTTGTG---GAACACGAAGCGGATTTTGGTTGGAAACCGCGGCGCTGTTAAAAA	771
Qy	260	-----AspLysLeuLeuProValGlyGluLeuLysIleArgGlyAlaHisAsnTyr	276
Db	772	CAAGGCNATGAAGATTGATTTCTACCAAGACATCCCGTGCAGGTTCGACAAAGCC	831
Qy	277	SerAsnAlaLeuAlaLeuAlaLeuGlyHisAlaValGlyLeuProPheAspAlaMet	296
Db	832	GCCAACTGTTATGCTGCGGTGCTTGTGCGAGGCGCGTGGTTGCGCGCGAAGCATTTG	891
Qy	297	LeuGlyAlaLeuLysAlaPheSerGlyLeuAlaHisArgCysGlnTrpValArgGluArg	316
Db	892	CTGGAAACACCTCAAAACCTTCCAAGGCTGCCCGACCGTGTGAAAAAATCGCGGAGAAA	951
Qy	317	GlnGlyValSerTyrTyrAspSerLysAlaThrAsnValGlyAlaAlaLeuAlaAla	336
Db	952	AACGGCGTGTATTATCATCGACGACAGCAAGGACGAGNATGTCGGCGGACCGCGCGCGG	1011
Qy	337	IleGluGlyLeuGlyAlaAspIleAspGlyLysLeuValLeuLeuAlaGlyGlyAspGly	356
Db	1012	ATTGCGCGTTG-----CAAAACCCGCTCTTCGTGATTTTGGCGCGCATGGCG	1059
Qy	357	LysGlyAlaAspPheHisAspLeuArgGluProValAlaAlaArgPheCysArgAlaValAl	376
Db	1060	AAAGGCGAGACTTCACGCCCTTCGCGCGCGCGTGTAAAGATAAGGCAAAAGCGGTTC	1119
Qy	377	LeuLeuGlyArgAspAlaGlyLeuIleAlaGlnAlaLeu---GlyAsnAlaValProLeu	395
Db	1120	CTGATCGCGCTCGATCGCGCGCAATCCGCGCGGATTTGGACGGCTCGCGCTTGAACCTG	1179
Qy	396	ValArgValAlaThrLeuAspGluAlaValArgGlnAlaAlaGluLeuAlaArgGluGly	415
Db	1180	ACCGACTGCGCTCACTTTGGAAGAGGCGGTTCAGACGGCATACGCCCAAGCGGAAGCGGCG	1239
Qy	416	AspAlaValLeuLeuSerProAlaCysAlaSerLeuAspMetPheLysAsnPheGluGlu	435
Db	1240	GATATTGCTCTGTCTACGCCCGCGCTGGCGAGTTCGATATATGTTTAAAGGCTACGCCAC	1299
Qy	436	ArgGlyArgLeuPheAlaLysAlaValGluGluLeu	447
Db	1300	CGTTCGGAAGTGTATTATCGAAGCGGTTTAAAGCGCTTTG	1335

## DEC 17 1955

AXX30137	ID	AAx30137 standard; DNA; 2193 BP.	QY	7 AspHisPheArg-----lleValValGlyLeuGlyLysSerGlyMetSerIleu	22
XX	XX				
XX	AC				
AXX30137;	Db		494	GATCAATTATAAATAAGGAAGTCCTGTGTTTAGTGGCCAGTCTGGTGAACTGC	553
XX	XX				
18-JUN-1999 (first entry)	Df		23	ValArgTyrLeuAlaArgGlyLeuPropheAlaValAlaAspThrArg-----	39
XX	XX				
Streptococcus pneumoniae MurD nucleotide sequence.	DE		554	GCTCGTTTTGTGGACAAGCTAGTGCCCATTTGTGACAGTAATATGGAAACCTTC	613
XX	XX				
Streptococcus pneumoniae; MurD; antibacterial; infection; otitis media;	KW		40	GluAsnProProGluLeuAlaThrLeuArgAlaGlnTyrProGlnValGluValArgCys	59
conjunctivitis; pneumonia; bacteraemia; sinusitis; pleural empyema;	KW		:::     :::		:::     :::
endocarditis; meningitis; ds.	KW		614	GACATCCAGCTGCCCAAAGTTTGTGGAA-----GAAGGATCAAGGTCAATTACA	664
XX	XX				
Streptococcus pneumoniae.	OS		60	GlyGlu-----LeuAspAlaGluPheLeuCysSerAlaArgGluLeuTyr	74
XX	XX				:::
EP906956-A2.	PN		665	GGTGGCCATCTTTTGGAACTCTTGGATGAAGAGTTTGCTCTTATGGTGAAA-	715
XX	XX				
07-APR-1999.	PD		75	ValSerProGlyLeuSerLeuArgThrProAlaLeuValGlnAlaAlaLysGlyVal	94
XX	XX		:	:::	:::
30-JUL-1998; 98EP-0306076.	Pf		716	---AATCCAGGTATCCCTCACAAATCCCATGATTGAAAGGGTTTGGCCAGGGAATT	772
XX	XX				
19-SEP-1997; 97US-0934481.	PR		95	ArgIleSerGlyAspIleAspLeuPheAlaArgGluAlaLysAlaProIleValaIatle	114
12-AUG-1997; 97US-0056335.	PR		:	:::	:::
(SMIK ) SMITHKLINE BEECHAM CORP.	PA		773	CCAGTCTTGACTGAGGTGGGAATTGGCTTATTGATTTCCAGAAGCACCGATTATTGGTATC	832
(SMIK ) SMITHKLINE BEECHAM PLC.	PA		115	ThrGlySerAsnAlaLysSerThrValThrLeuValGlyGluMetAlaValAlaala	134
XX	XX				
XX	XX				

Db 833 ACAGGATCGACGGTAAGACACACACAGACTATGATTGGGAACTTTTGTACTGCTGCT 892  
QY 135 AspLysArgValAlaValGlyGlyAsnLeuGlyThrProAlaLeuAspLeu-----Leu 152  
Db 893 GGGCAACATGGTCTTTTATCAGGAATATCGGCTATCCAGCTAGTCAAGTGGCCCAACT 952  
QY 153 AlaAspIleLeuLeuTyrValLeuGluLeuSerPheGlnLeuGluThrCysAsp 172  
Db 953 GCGTCGACAGACACGCTGTTGTAAGTCTTCTTCCAACTCATGGGTGTTCAA 1012  
QY 173 ArgLeuAsnAlaGluValAlaThrValLeuAsnValSerGluAspHisMetAspArgTyr 192  
Db 1013 GAATTCATCCAGAGATTCGGTTATACCACTTATGCCAACTCATATCGACTACCAT 1072  
QY 193 AspGlyMetAlaAspTyrHisLeuAlaLysHisArgIle-----PheArgGlyAla 209  
Db 1073 GGGTCATTTTCGAATATAGCAGCAAGTGAATATCCAGAACAGATGACAGCAGCT 1132  
QY 210 ArgGlnValValAlaAsnArgAlaAspAlaLeuThrArgProLeuIleAlaAspThrVal 229  
Db 1133 GATTCCTGTGCTTGAACCTTAAATCAAGACTTGACTTCCAGACAGAACCCACTGTTGTA 1192  
QY 230 ProCysTrpSerPheGlyLeuAsnLysProAspPheLysAlaPheGlyLeuIleGluGlu 249  
Db 1193 CCATTTTCAACA-----CTTGAAAAGTTGAT-----GGAGCTTATCTGGAA 1234  
QY 250 AspGlyGlnLysTrpLeuAlaPheGlnPheAspLysLeuLeuProValGlyGluLeuLys 269  
Db 1235 GATGTGCAA-----CTCTACTTCCGTGGTGAAGTAGTATCGCAGCGAATGAAATCGGT 1288  
QY 270 IleArgGlyAlaHisAsnTyrSerAsnAlaLeuAlaAlaLeuAlaLeuGlyHisAlaVal 289  
Db 1289 GTTCCAGGTAGCCACAAATGTGAAATGCCCTGGCAGATTGCTGTAGCCCAAGCTTCGT 1348  
QY 290 GlyLeuProPheAspAlaMetLeuGlyAlaLeuLysAlaPheSerGlyLeuAlaHisArg 309  
Db 1349 GGTGTGGACAATCAACCATCAAGGAACCTCTTCAGCCTTCGGTGGTCAACACCGT 1408  
QY 310 CysGlnTrpValArgGluArgGlnGlyValSerTyrAspSerLysAlaThrAsn 329  
Db 1409 CTCGATTTGTGGATGACATCAAGGCTGTAATCTATACGACAGTAAATCAACTAAT 1468  
QY 330 ValGlyAlaAlaLeuAlaIleGluGlyLeuGlyAlaAspIleAspGlyLysLeuVal 349  
Db 1469 ATCTGGCTACTCAAAAGCCCTGTGAGGATTTGAC-----AACAGCAAGGTCGTC 1519  
QY 350 LeuLeuAlaGlyGlyAspGlyLysGlyAlaAspPheHisAspLeuArgGluProValAla 369  
Db 1520 TTGATTCAGGTGGTTGGACCGTGGCAATGAGTTGACGAATTG---GTGCCACACATT 1576  
QY 370 ArgPheCysArgAlaValValLeuLeuGlyArgAspAlaGlyLeuIleAlaGlnAlaLeu 389  
Db 1577 ACTGGACTCAGAAGATGTCATCTCGGTCAATCTGCAGAACGTGTCAACGGGCAGCA 1636  
QY 390 GlyAsnAla---ValProLeuValArgValAlaThrLeuAspGluAlaValArgGlnAla 408  
Db 1637 GACAAGGCTGGTGTGCTTATGTGGAGCGCACAGATATTGCAGATCGCACCGCAAGGCC 1696  
QY 409 AlaGluLeuAlaArgGlyAlaValLeuLeuSerProAlaCysAlaSerLeuAsp 428  
Db 1697 TATGAGCTGCGACTCAGGAGATGTGTTCTTCTTAGTCTGCCAATGCCAGCTGGGAT 1756  
QY 429 MetPheLysAsnPheGluGluArgGlyArgLeuPheAlaLysAlaValGluGluLeu 447  
Db 1757 ATGTATGCTAACTTTGAAGTAGCTAGCGACCTCTTTTATCGACACAGTAGCGGAGTTA 1813

Search completed: August 14, 2003, 07:47:54  
Job time : 2055 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n.model

Run on: August 14, 2003, 06:15:45 ; Search time 4239 Seconds  
(without alignments)  
4323.546 Million cell updates/sec

Title: US-09-701-229-2

Perfect score: 2260

Sequence: 1 MSLIASDHFRIVVGLGSGM.....MFKNFEGRGLFAKAVEELA 448

Scoring table: BLOSUMP62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 2888711 seqs, 20454813386 residues

Total number of hits satisfying chosen parameters: 4991172

Minimum DB-seq length: 25  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL=frame+p2n.model -DRV=xlp  
-O=/cgn2\_1/USPTO.spool/US09701229/runat\_09082003\_161157\_4380/app\_query.fasta\_1.647  
-DB=GenEmbl -QFMT=fastpct -SUFFIX=p2n.rge -MINMATCH=0.1 -LOOPEL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blosom62 -TRANS=human40.cdd -LIST=45  
-DOALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM-ext -HEAPSIZE=500 -MINLEN=25 -MAXLEN=2000000000  
-USER=US09701229.ecgn\_1\_1.5265,runat\_09082003\_161157\_4380 -NCPU=6 -ICPU=3  
-NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOPOP=6  
-FGAPEXT=7 -FGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : GenEmbl :  
1: gb\_ba :  
2: gb\_hgt :  
3: gb\_in :  
4: gb\_ov :  
5: gb\_ov :  
6: gb\_pat :  
7: gb\_ph :  
8: gb\_pl :  
9: gb\_pr :  
10: gb\_ro :  
11: gb\_sts :  
12: gb\_sy :  
13: gb\_un :  
14: gb\_vi :  
15: em\_ba :  
16: em\_fun :  
17: em\_hum :  
18: em\_in :  
19: em\_mu :  
20: em\_om :  
21: em\_or :  
22: em\_ov :  
23: em\_pat :  
24: em\_ph :  
25: em\_pl :  
26: em\_ro :  
27: em\_sts :  
28: em\_un :

29: em\_vi :  
30: em\_hgt\_hum :  
31: em\_hgt\_inv :  
32: em\_hgt\_other :  
33: em\_hgt\_mus :  
34: em\_hgt\_pin :  
35: em\_hgt\_rod :  
36: em\_hgt\_mam :  
37: em\_hgt\_vrt :  
38: em\_sy :  
39: em\_hgt\_hum :  
40: em\_hgt\_mus :  
41: em\_hgt\_other :\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
c 1	2260	100.0	24000	1	AE004856	AE004856 Pseudomon
2	2249	99.5	5280	1	AY008276	AY008276 Pseudomon
3	1835	81.2	301995	1	AE016779	AE016779 Pseudomon
c 4	1829	80.9	311600	1	AE016871	AE016871 Pseudomon
5	1023	45.3	301727	1	AE016960	AE016960 Cxiella
6	990	43.8	251050	1	AL627265	AL627265 Salmonell
7	990	43.8	300169	1	AE016834	AE016834 Salmonell
8	987	43.7	22348	1	AE008699	AE008699 Salmonell
9	986	43.6	2608	1	ECMUROY	X51584 Escherichia
10	986	43.6	21757	1	AE000118	AE000118 Escherich
11	986	43.6	28277	6	AX191720	AX191720 Sequence
12	986	43.6	28277	15	EC2MIN	X55034 E. coli 2 m
13	986	43.6	111408	1	EC0110K	D10483 Escherichia
14	985.5	43.6	1317	6	AX189058	AX189058 Sequence
15	984	43.5	12791	1	AE005185	AE005185 Escherich
16	984	43.5	281530	1	AP002550	AP002550 Escherich
17	984	43.5	300409	1	AE016755	AE016755 Escherich
18	978	43.3	11441	1	AE015046	AE015046 Shigella
19	973	43.1	290029	1	AE016978	AE016978 Shigella
20	972	43.0	1629	1	ECMURD	X17609 Escherichia
c 21	968	42.8	13071	1	AE015855	AE015855 Shewanell
c 22	965.5	42.7	11726	1	AE013965	AE013965 Yersinia
c 23	965.5	42.7	201050	1	AJ414143	AJ414143 Yersinia
24	927	41.0	316050	1	BX321859	BX321859 Nitrosomo
25	908	40.2	21000	1	AB052554	AB052554 Shewanell
c 26	902	39.9	248308	2	AC073937	AC073937 Mus muscu
c 27	889.5	39.4	96109	6	AX067460	AX067460 Sequence
28	879	38.9	11498	1	U32793	U32793 Haemophilus
29	879	38.9	11498	6	AX191762	AX191762 Sequence
30	879	38.9	110000	6	AR274513_11	Continuation (12 o
31	879	38.9	110000	6	AR274513_12	Continuation (13 o
32	875.5	38.7	298900	1	AP005074	AP005074 Vibrio pa
c 33	864	38.2	301442	1	AE016798	AE016798 Vibrio vu
c 34	859	38.0	188050	1	AL646072	AL646072 Ralstonia
c 35	844.5	37.4	12232	1	AE006048	AE006048 Pasteurel
c 36	832	36.8	22201	1	AE004310	AE004310 Vibrio ch
c 37	757.5	33.5	348600	1	AB063521	AB063521 Wigglewo
c 38	738	32.7	259498	2	AC020876	AC020876 Mus muscu
c 39	716	31.7	110000	6	BD061520_2	Continuation (3 of
c 40	716	31.7	347550	1	AP001118	AP001118 Buchnera
c 41	710	31.4	195767	1	NMA722491	AL162758 Neisseria
c 42	704	31.2	10629	1	AE002398	AE002398 Neisseria
43	704	31.2	349980	6	AX044029	AX044029 Sequence
c 44	694	30.7	310003	1	AE014016	AE014016 Buchnera
c 45	693.5	30.7	11683	1	AE013119	AE013119 Thermoana

#### ALIGNMENTS

RESULT 1

AE004856/c	AE004856	24000 bp	DNA	linear.	BCT 19-FEB-2003								
LOCUS	Pseudomonas aeruginosa PA01, section 417 of 529 of the complete genome.												
DEFINITION	Pseudomonas aeruginosa PA01, section 417 of 529 of the complete genome.												
ACCESSION	AE004856	AE004091											
VERSION	AE004856.1	GI:9950633											
KEYWORDS													
SOURCE	Pseudomonas aeruginosa PA01												
ORGANISM	Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales; Pseudomonadaceae; Pseudomonas.												
REFERENCE	1. (bases 1 to 24000)												
AUTHORS	Stover,C.K., Pham,X.-Q.T., Erwin,A.L., Mizoguchi,S.D., Warrenner,P., Hickey,M.J., Brinkman,F.S.I., Hufnagle,W.O., Kowalik,D.J., Lagrou,M., Garber,R.L., Goltry,L., Tolentino,E., Westbrook-Wadman,S., Yuan,Y., Brody,L.L., Coulter,S.N., Folger,K.R., Kas,A., Larbig,K., Lim,R.M., Smith,K.A., Spencer,D.H., Wong,G.K.-S., Wu,Z., Paulsen,I.T., Reizer,J., Sater,M.H., Hancock,R.E.W., Lory,S. and Olson,M.V.												
TITLE	Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pathogen												
JOURNAL	Nature 406 (6799), 959-964 (2000)												
MEDLINE	20437337												
PUBMED	10984043												
REFERENCE	2. (bases 1 to 24000)												
AUTHORS	Stover,C.K., Pham,X.-Q.T., Erwin,A.L., Mizoguchi,S.D., Warrenner,P., Hickey,M.J., Brinkman,F.S.I., Hufnagle,W.O., Kowalik,D.J., Lagrou,M., Garber,R.L., Goltry,L., Tolentino,E., Westbrook-Wadman,S., Yuan,Y., Brody,L.L., Coulter,S.N., Folger,K.R., Kas,A., Larbig,K., Lim,R.M., Smith,K.A., Spencer,D.H., Wong,G.K.-S., Wu,Z., Paulsen,I.T., Reizer,J., Sater,M.H., Hancock,R.E.W., Lory,S. and Olson,M.V.												
TITLE	Direct Submission												
JOURNAL	Submitted (16-MAY-2000) Department of Medicine and Genetics, University of Washington Genome Center, University of Washington, Box 352145, Seattle, WA 98195, USA												
REFERENCE	3. (bases 1 to 24000)												
AUTHORS	Pseudomonas aeruginosa Community Annotation Project (PseudoCAP)												
CONSKTM	Direct Submission												
TITLE	Submitted (04-FEB-2003) Department of Molecular Biology and Biochemistry, Simon Fraser University, 8888 University Dr., Burnaby, British Columbia V5A 1S6, Canada												
JOURNAL													
COMMENT	This represents the February 3, 2003 version of the continually updated, reviewed, Pseudomonas aeruginosa PA01 genome annotation, from PseudoCAP (see <a href="http://www.pseudomonas.com">http://www.pseudomonas.com</a> for latest updates and links to alternate annotations). PseudoCAP is coordinated by Fiona S.L. Brinkman (Simon Fraser University, Canada) and Robert E.W. Hancock (University of British Columbia, Canada). We welcome submission through <a href="http://www.pseudomonas.com">www.pseudomonas.com</a> of any proposed changes.												
'Protein name confidence' is used to rate our confidence of the accuracy of the protein name.													
Class 1: Function experimentally demonstrated in P. aeruginosa.													
Class 2: Function of highly similar gene experimentally demonstrated in another organism (and gene context consistent in terms of pathways its involved in, if known).													
Class 3: Function proposed based on presence of conserved amino acid motif, structural feature or limited sequence similarity to an experimentally studied gene.													
Class 4: Homologs of previously reported genes of unknown function, or no similarity to any previously reported sequences.													
-----													
FEATURES	Location/Qualifiers												
Source	1..24000												
/organism="Pseudomonas aeruginosa PA01"													
/mol_type="genomic DNA"													
/strain="PA01"													
/db_xref="taxon:208964"													
complement(91..1308)													
/gene="argJ"													
gene													
complement(91..1308)													
CDS	locus_tag="PA4402"												
-----													
/gene="argJ"													
/locus_tag="PA4402"													
/note="Protein name confidence: Class 4 (homologs of													
to functionally studied protein); Subcellular													
localization: Cytoplasmic; Subcellular localization													
confidence: Class 2; Alt protein name: ornithine													
transacetylase"													
/codon_start=1													
/transl_table=11													
/product="glutamate N-acetyltransferase"													



previously reported genes of unknown function, or no similarity to any previously reported sequences")

```
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AA007793.1"
/db_xref="GI:9950637"
/translation="MLFAAQRICRLQQLLESLOPAARPYCHVASWRGGCLILLIVTD
GHWATRLVQORLQRLQAFEFANLILFKVQPTTSTQRAARSLTILSAVAENL
RESAEIVGNPRLEALERLASRAEKSQE"
/complement(5865..6776)
/gene="lpxc"
/locus_tag="PA4406"
/notes="synonyms: envA, asmB"
/complement(5865..6776)
/gene="lpxC"
/locus_tag="PA4406"
/notes="Protein name confidence: Class 1 (function experimentally demonstrated in P. aeruginosa); Alt protein name: EnvA protein"
/codon_start=1
```

gene

CDS

```
/transl_table=11
/product="UDP-3-O-acetyl-N-acetylglucosamine deacetylase"
/protein_id="AA007794.1"
/db_xref="GI:9950638"
/translation="MIKQRTLNIRATGVGLHSGEKVYTLKPAPVDTGIVFCRTDL
DPVEIPARAENYGETMTSTLVKGDVDTVBHLLSAMAGLIDINAYVELSASEVPI
MDGSGPFVFLIQSAGLOEOEAAKFKIRIKREVSVEGDKRAVFPDFGPKVSEFIDF
DHPVFRGTQOASVDPSSTSEVKEVSARTFGPMRDIEYLRSONLALGSVENAIVVD
ENRVLEDGLRYDEDFVKHILDAIDGLILLGSLGIFGFGKSHALANQLLRTLIA
DKDAWEVTFEDARTAPISYMRPAAV"
/complement(6889..8073)
/gene="ftsZ"
/locus_tag="PA4407"
/complement(6889..8073)
/gene="ftsZ"
/locus_tag="PA4407"
/notes="Protein name confidence: Class 2 (high similarity to functionally studied protein)"
/codon_start=1
```

gene

CDS

```
/transl_table=11
/product="cell division protein FtsZ"
/protein_id="AA007795.1"
/db_xref="GI:9950639"
/translation="MFELVDNIAQTAVIKIVGYGGGGNVAHNMKNVVEGFICAN
TDAQLKNTAARTVLQPGVTKGLGAGANPEVGRQAALDRERISEVLEGADMVIT
TGMGGGTGGAAPILIAEVAKEGILTVAVVTRPFPFEGKRMOIADGIRALASVDS
LITIPNEKLLTILGKASLLAAFAKADDVLGAVRGISDIIKPPGMINVDFADVKTM
SEMGMMGTGASGNRAREATEAARNPLEDVNLQAGILVNIATGDSLSEY
SDVGNIEGFASHEATVKGTVIDAMRDELHVTVVATGLGARLEKPKVVDNTVOGS
AAQAAAQAQREQQSVNYRDLRPTVMRNSHGSAAATAAKLINPQDDLDYLDIPAFLRQ
AD"
/complement(8124..9377)
/gene="ftsA"
/locus_tag="PA4408"
/complement(8124..9377)
```

gene

## Alignment Scores:

```
Pred. No.: 7.91e-138 Length: 24000
Score: 2260.00 Matches: 448
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 1 Gaps: 0
```

US-09-701-229-2 (1-448) x AE004856 (1-24000)

```
Qy 1 MetSerLeuIleAlaSerAspHisPheArgIleValValGlyLeuGlyLysSerGlyMet 20
Db 16265 ATGAGCCTGATCCGCTCCGACCACTTCCCGATCGTGTTCGCGCCTCGGCAAGCGGCATG 16206
Qy 21 SerLeuValArgTyrLeuAlaArgGlyLeuProPheAlaValValAlaAspThrArgGlu 40
Db 16205 TCCTGTGTCGCTACCTGGCGCGCGCGGCTTCCCTTTCGCGCTGTGTCGCGCCTCGGCAAGCGGCATG 16146
```

```
Qy 41 AsnProProGluLeuAlaThrLeuArgAlaGlnTyrProGlnValGluValAlaArgCysGly 60
Db 16145 AACCCGCGGAGCTGGCCACCCCTGCGTGCACCATATCCGAGGTGGAAGTGCCTGGCGG 16086
Qy 61 GluLeuAspAlaGluPheLeuCysSerAlaArgGluLeuTyrValSerProGlyLeuSer 80
Db 16085 GAACTCGACCGGAGTTCTCTGTCGCGCGGCAACTCTATGTGACGCCCGGCTTCG 16026
Qy 81 LeuArgThrProAlaLeuValGlnAlaAlaAlaLysGlyValArgIleSerGlyAspIle 100
Db 16025 CTGCGCACCCCTGCGTGTGACAGCGCGCGGAAAGCGGTGCGCATCTCCGTTGACATC 15966
Qy 101 AspLeuPheAlaArgGluAlaLysAlaProIleValAlaIleThrGlySerAsnAlaLys 120
Db 15965 GATCTCTTCGCGCGGAGCGAAGGCCCGCATGTCGCCATCACCGGTTCCACGCCGAG 15906
Qy 121 SerThrValThrThrLeuValGlyGluMetAlaValAlaAlaAspLysArgValAlaVal 140
Db 15905 AGCACCGTGACACCCCTGTTGGCGAAATGCGGTGGCGGACAAAGCGTGTGCGCGTC 15846
Qy 141 GlyGlyAsnLeuGlyThrProAlaLeuAspLeuAlaAspIleGluLeuTyrVal 160
Db 15845 GCGCGCAACCTCGGCGCGCTCGACCTGCTGCGCGCGACGACATCGAGCTGACGTG 15786
Qy 161 LeuGluLeuSerSerPheGlnLeuGluThrCysAspArgLeuAsnAlaGluValAlaThr 180
Db 15785 TTGGAGCTGTCGAGCTTCCAGCTGGAACCTGCGATCGCTCAACGCCGAGGTGGCGACC 15726
Qy 181 ValLeuAsnValSerGluAspHisMetAspArgTyrAspGlyMetAlaAspTyrHisLeu 200
Db 15725 GTGCTGAACGTCACGCAAGACCATATGATGATCGTACGACGCGATGGCTGACTACACCTG 15666
Qy 201 AlaLysHisArgIlePheArgGlyAlaArgGluValValValAsnArgAlaAspAlaLeu 220
Db 15665 GCCAAGCACCGGATCTTCCGCGGTGGCGCGCGAGTGTGTGTAATCCGCGGATGGCGCTG 15606
Qy 221 ThrArgProLeuIleAlaAspThrValProCysTrpSerPheGlyLeuAsnLysProAsp 240
Db 15605 ACCCGACCGCTGATCGCGCATACCGTCCGCTGCTGCTGCTGCGCTTCCGCTTCAACGCCGAC 15546
Qy 241 PheLysAlaPheGlyLeuIleGluAspGlyGlnLysTrpLeuAlaPheGlnPheAsp 260
Db 15545 TTCAAGCTTTCGCGCTGATCGAGGAAGCGGCGAGGTGGCTTCCAGTTCGAC 15486
Qy 261 LysLeuLeuProValGlyGluLeuLysIleArgGlyAlaHisAsnTyrSerAsnAlaLeu 280
Db 15485 AAGCTGCTGCGGTGGCGAAGTCAAGATCGTGGCGCGCCACCACTATTCCAACGCCGCTC 15426
Qy 281 AlaAlaLeuAlaLeuGlyHisAlaValGlyLeuProPheAspAlaMetLeuGlyAlaLeu 300
Db 15425 GCGCGCTGGCGTGGGCCATCGGTGCGCTGCGCTGCGCTTCCGCGCATGCTCGCGCGCTG 15366
Qy 301 LysAlaPheSerGlyLeuAlaHisArgCysGlnTrpValArgGluArgGlnGlyValSer 320
Db 15365 AAGCGTTCGCGCTTTCGCGCTGATCGCTGCGAGTGGTATCGAGCGGCGGCGGCTGAGC 15306
Qy 321 TyrTyrAspAspSerLysAlaThrAsnValGlyAlaAlaLeuAlaIleGluGlyLeu 340
Db 15305 TACTACGACGATTCGAAGGCCACCAAGCTGCGCGCGCCCTTGGCGGCGATCGAGGGGCTG 15246
Qy 341 GlyAlaAspIleAspGlyLysLeuValLeuLeuAlaGlyGlyAspGlyLysGlyAlaAsp 360
Db 15245 GGTGCCGACATCGACGCAAGCTGCTGCTCGCGCGGAGAGCGCAAGGGCGCGAT 15186
Qy 361 PheHisAspLeuArgGluProValAlaArgPheCysArgAlaValValLeuLeuGlyArg 380
Db 15185 TTCATGACCTGCGGAGCGGTCGCGCTTCTGCGCGGCGGTGGTACTGCTTGGCGCT 15126
Qy 381 AspAlaGlyLeuIleAlaGlnAlaLeuGlyAsnAlaValProLeuValArgValAlaThr 400
Db 15125 GACGCCGCGCTGATTTGCCCGAGGCACTGGGCAACGCGGTACCGCTGCTGCGGCTCGCAACG 15066
```



```
Db 3929 ATGAGCGTGATCGCCCTCCGACCACTTCCCGCATCGTTGTCCGCTCGGCAAGCGCGCATG 3988
Qy 21 SerLeuValArgTyrLeuAlaArgGlyLeuProPheAlaValValAspThrArgGlu 40
Db 3989 TCCTGGTGGCTACCTGGCGCGCGGGCTTCCCTTTCGCCGTGGTGCATACCCGAGAG 4048
Qy 41 AspProGluLeuAlaThrLeuArgAlaGlnTyrProGlnValGluValArgCysGly 60
Db 4049 AACCCGCGGAGCTGGCTACCTCCGTCGCGTCCGACATATCCGAGGTGGAAGCGCGTTCGGC 4108
Qy 61 GluLeuAspAlaGluPheLeuCysSerAlaArgGluLeuTyrValSerProGlyLeuSer 80
Db 4109 GAATCTGACCCGAGTTCCTCTCTCCGCGCGGAATCTATGTACGCGCGGCTTGTGTCG 4168
Qy 81 LeuArgThrProAlaLeuValGlnAlaAlaLysGlyValArgIleSerGlyAspIle 100
Db 4169 CTGCGCACCCCTCGCTGTGTACAGCGCGCGCGCAAGGCGTGCATCTCCGCTGACATC 4228
Qy 101 AspLeuPheAlaArgGluAlaLysAlaProIleValAlaIleThrGlySerAsnAlaLys 120
Db 4229 GATCTCTTCGCGCGGAGGCGAAGCCCGCATCGTCCCATCACCGGTTCCAAACGGGAAG 4288
Qy 121 SerThrValThrLeuValGlyGluMetAlaValAlaAlaAspLysArgValAlaVal 140
Db 4289 AGCACCTGACCACTTCGTTGGCGAATGCGGTGGCGCGGACAGCGGTGCGCGCTG 4348
Qy 141 GlyGlyAsnLeuGlyThrProAlaLeuAspLeuAlaAspAspIleGluLeuTyrVal 160
Db 4349 GCGGGAACCTCGCACCCCGCGCTCGACCTGCTGCGCGACACATCGAGCTGACGTG 4408
Qy 161 LeuGluLeuSerSerPheGlnLeuGluThrCysAspArgLeuAsnAlaGluValAlaThr 180
Db 4409 TTGGAGCTGTCGAGCTTCCAGCTGGAACCTGCGCATCGCTCAACGCGAGGTGGCGACC 4468
Qy 181 ValLeuAsnValSerGluAspHisMetAspArgTyrAspGlyMetAlaAspTyrHisLeu 200
Db 4469 GTGCTGACGCTCAGGACGACCATATGATCGGTACGACGCGATGGCTGACTACCACTG 4528
Qy 201 AlaLysHisArgIlePheArgGlyAlaArgGlnValValValAsnArgAlaAspAlaLeu 220
Db 4529 GCCAAGCAGCGGATCTTCCGCGTGGCGCGGATGCTGTGTAATCGCGCGCATGCGCTG 4588
Qy 221 ThrArgProLeuIleAlaAspThrValProCysTrpSerPheGlyLeuAsnLysProAsp 240
Db 4589 ACCGACGCTGATCGCGCATACCGTCCGCTGTGTCGTCGCGCTGGAACAGCGCGGAC 4648
Qy 241 PheLysAlaPheGlyLeuIleGluAspGlyGlnLysTrpLeuAlaPheGlnPheAsp 260
Db 4649 TTCAAGCTTTCGCGCTGATCGAGGAGACGGCGAGTGGTGGCTGGCTTCAGATTCCGCG 4708
Qy 261 LysLeuLeuProValGlyGluLeuLysIleArgGlyAlaHisAsnTyrSerAsnAlaLeu 280
Db 4709 AAGCTGCTGCGGTTGGCAACTGAAGATCCGTGGCGGCCCCACAACTATTTCACAGCGCTC 4768
Qy 281 AlaAlaLeuAlaLeuGlyHisAlaValGlyLeuProPheAspAlaMetLeuGlyAlaLeu 300
Db 4769 GCGCGCTGCGCTGGGCCCATCGGTGCGCTGCGCTGTCGACGCGCATGCTCGCGCGCGCTG 4828
Qy 301 LysAlaPheSerGlyLeuAlaHisArgCysGlnTrpValArgGluArgGlnGlyValSer 320
Db 4829 AAGCGCTTTTCGCGCTGGCTCATCGCTGCGCAGTGGGTACGCGAGCGGCGGCGGTGAGC 4888
Qy 321 TyrTyrAspSerLysAlaThrAsnValGlyAlaAlaLeuAlaIleGluGlyLeu 340
Db 4889 TACTACGACCATTCAGAGCCACCAAGCTCGGCGCGCCCTCGCGGCGATCGAGGGGCTG 4948
Qy 341 GlyAlaAspIleAspGlyLysLeuValLeuAlaGlyAspGlyLysGlyAlaAsp 360
Db 4949 GGTGCCACATCGACGCAAGCTGGTGTCTGCTCGCGCGGAGAGCGCAAGGCGCGCAT 5008
Qy 361 PheHisAspLeuArgGluProValAlaArgPheCysArgAlaValValLeuLeuGlyArg 380
Db 5009 TTCATGACCTGCGGAGCGCGTTCGCGCGCTTCTGCGGCGGTGGTGGTACTGCTTGGCGCGT 5068
```

---

```
Qy 381 AspAlaGlyLeuIleAlaGlnAlaLeuGlyAsnAlaValProLeuValArgValAlaThr 400
Db 5069 GACGCGGGCTGATTGCCAGGCACCTGGGCAACGGGTACCGCTGGTGGCGGTGCGCAACG 5128
Qy 401 LeuAspGluAlaValArgGlnAlaAlaGluLeuAlaArgGluGlyAspAlaValLeuLeu 420
Db 5129 CTGGAGAGACAGTCGCGGAGCGCGGAGCTGGCGCGGAGGCGGATGGCGGTGCTGTTG 5188
Qy 421 SerProAlaCysAlaSerLeuAspMetPheLysAsnPheGluGluArgGlyArgLeuPhe 440
Db 5189 TCGCCGCGCTGCGGAGCGCTGGACATCTTCAAGACTTCGAGAACCGGACGCGCTGTTTC 5248
Qy 441 AlaLysAlaValGluGluLeuAla 448
Db 5249 GCCAAGCGGTAGAGGAGCTAGCG 5272

RESULT 3
AE016779 301995 bp DNA linear BCT 12-DEC-2002
LOCUS Pseudomonas putida KT2440 section 6 of 21 of the complete genome.
DEFINITION AE016779 AE015451
ACCESSION AE016779.1 GI:26557023
VERSION
KEYWORDS
SOURCE
ORGANISM Pseudomonas putida KT2440
Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
Pseudomonadaceae; Pseudomonas.
REFERENCE
1 (bases 1 to 301995)
Nelson, K., Paulsen, I., Weinl, C., Dodson, R., Hilbert, H., Fouts, D.,
Gill, S., Pop, M., Martins Dos Santos, V., Holmes, M., Brinkac, L.,
Beanan, M., DeBoy, R., Daugherty, S., Kolonay, J., Madupu, R.,
Nelson, W., White, O., Peterson, J., Khouri, H., Hance, I., Lee, P.,
Holtzapple, E., Scanlan, D., Tran, K., Moazzaz, A., Utterback, T.,
Rizzo, M., Lee, K., Kosack, D., Moestl, D., Wedler, H., Lauber, J.,
Hoheisel, J., Straetz, M., Helm, S., Kiewitz, C., Eisen, J., Timmis, K.,
Duesterhoft, A., Tummeler, B. and Fraser, C.
Complete genome sequence and comparative analysis of the
metabolically versatile Pseudomonas putida KT2440
Environ. Microbiol. 4 (12), 799-808 (2002)
2 (bases 1 to 301995)
Nelson, K., Paulsen, I., Weinl, C., Dodson, R., Hilbert, H., Fouts, D.,
Gill, S., Pop, M., Martins Dos Santos, V., Holmes, M., Brinkac, L.,
Beanan, M., DeBoy, R., Daugherty, S., Kolonay, J., Madupu, R.,
Nelson, W., White, O., Peterson, J., Khouri, H., Hance, I., Lee, P.,
Holtzapple, E., Scanlan, D., Tran, K., Moazzaz, A., Utterback, T.,
Rizzo, M., Lee, K., Kosack, D., Moestl, D., Wedler, H., Lauber, J.,
Hoheisel, J., Straetz, M., Helm, S., Kiewitz, C., Eisen, J., Timmis, K.,
Duesterhoft, A., Tummeler, B. and Fraser, C.
Direct Submission
Submitted (05-NOV-2002) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA
JOURNAL
FEATURES
source
1. 301995
/organism="Pseudomonas putida KT2440"
/mol_type="genomic DNA"
/strain="KT2440"
/db_xref="taxon:160488"
89. .712
/gene="sspA"
/locus_tag="Pp1320"
89. .712
/gene="sspA"
/locus_tag="Pp1320"
/notes="similar to GP:9864180; identified by sequence
similarity; putative"
/transl_table=11
/product="stringent starvation protein A"
/protein_id="AAG66944.1"
/db_xref="GI:24982777"
/translation="MGATNRLACYSDPADHYSHRVLRLVLAEKGVSVOLIDVDPAHLPR
KLAENPYGSVPTLVDRDLALYESTVVMYLEERYHPPLPMPYPVARGNSRLMHRI
```

QRDWCALADTVLDPRSSAARTARKALRESITGVSLFSEFACFMSEQSLVDCCLL  
 PTLWRLPVIGIELPQAKPLLDYMERQFAPEFQASLSVERMKL"  
 693. .1154  
 /gene="sspb"  
 /locus\_tag="pp1321"  
 693. .1154  
 /gene="sspb"  
 /locus\_tag="pp1321"  
 /note="similar to GB:S50157, GB:X60221, SP:P24539, and  
 PID:509291; identified by sequence similarity; putative"  
 /codon\_start=1  
 /transl\_table=11  
 /product="stringent starvation protein B"  
 /protein\_id="AAN66945.1"  
 /db\_xref="GI:24982778"  
 /translation="MKCAFKEPLMNSRRPYLYRALLYEIVDNDCTPHMLYNAEPKV  
 QVPFASDQGVILNISAVNSLHMDNDVSGFSGVAHSLFVPGAILGIYARE  
 NQGMVFEEFSEMDGDEDDVQPPDDGPPDGGQPPRTGPRSLKVR"  
 complement(1216. .1794)  
 /locus\_tag="pp1322"  
 complement(1216. .1794)  
 /locus\_tag="pp1322"  
 /codon\_start=1  
 /transl\_table=11  
 /product="lipoprotein, putative"  
 /protein\_id="AAN66946.1"  
 /db\_xref="GI:24982779"  
 /translation="WTPMRLGLMALTCLSVTGCSSVLTSTRNSPIEDDCTRTIGSK  
 IDDSLIETKAAVNISKANPNLDKGSIVVSYNGIYLLAGQTPRADLSLAQTSQV  
 QRKXVHNLQVMPQSSILARNNDLWLTIKTKTQMLTDSAVPSSRIKIVTENGIVYLL  
 GLVTQKANSATAVQGVSGQIKVLFYID"  
 complement(1791. .2378)  
 /gene="gmha"  
 /locus\_tag="pp1323"  
 complement(1791. .2378)  
 /gene="gmha"  
 /locus\_tag="pp1323"  
 /note="identified by match to PFAM protein family HMM  
 PF01380"  
 /codon\_start=1  
 /transl\_table=11  
 /product="phosphoheptose isomerase"  
 /protein\_id="AAN66947.1"  
 /db\_xref="GI:24982780"  
 /translation="MOSRIIRLFASIEIKQAMDILAPHIEQASLVNVALNEGKM  
 LACNGSGAGQAFHSSEILNFRERESLPALITDTSSTLTANDYSYNEVFSKQ  
 IRLVQPGDVLIAITSGNSANVIOAIOAHDRMIVVALTGRDGGNASLLPEDVE  
 IRVPSVTWARTQEVHLLAHCLDLDLSQLFGSE"  
 complement(2471. .2845)  
 /locus\_tag="pp1324"  
 complement(2471. .2845)  
 /locus\_tag="pp1324"  
 /note="identified by match to TIGR protein family HMM  
 TIGR00252"  
 /codon\_start=1  
 /transl\_table=11  
 /product="conserved hypothetical protein TIGR00252"  
 /protein\_id="AAN66948.1"  
 /db\_xref="GI:24982782"  
 /translation="MMAASPTRAGQAAETQALEVLOGIGLQLLARNWRCKGGGLDLV  
 MLDATVVFVSVRYLHAGFGGALDSIDGRQKRLVLAATLFLQKEAHGWNPCRFDV  
 VALQSHHAGRLQWLKNAFEC"  
 complement(2842. .4659)  
 /locus\_tag="pp1325"  
 complement(2842. .4659)  
 /locus\_tag="pp1325"  
 /codon\_start=1  
 /transl\_table=11  
 /product="lipoprotein, putative"  
 /protein\_id="AAN66949.1"  
 /db\_xref="GI:24982783"  
 /translation="WIACLRLLTALCLLAALAACASSPSSSILGELPRTPDASIQELLE  
 KAASSKSNEDALLRLSADLAYKQDFPRAARILEQVPLDTLPAQOVFASTLADEL

AMSNRKPAALTALAHPSLQRLAEPLADLQARTYSVHAALAEADQALAAQAQRILL  
 PLLGQAASNTNDALWALVASLPABQLQOQNTDQTLAQWTSAPAVKAGTLEQOQAA  
 IEAVVKQHPDHPAAQQLPLATLTKELASQPLTKALLPQEGPLAGVARGLRDQFMA  
 AHFQOQSGQPAPAVQVDESSRIGSLDDFYRQAQAGVQVLTGPLEKFLYKQIAAKPQ  
 LPTITLALNADAGQKAPQOLFQGLAAEDAREVARRARADVRAVALVPSGEMGD  
 RVLAAFRODWEGNGTLLAARIQAPVALAQIADFLQRQSEGRAKLSQTVGSGIA  
 AQPSSRDIDFILTASTPQQAQKIKPTLNFQVAGDVPIYATSNLYSASGDILOYNDMN  
 GIRFCETPMLDITNSLRQOVQOOWPOAAGSILGRLYANGVDAYS LAPRLGQALKALPDN  
 RVLGSLGSLINANQVRERQLPWAEEFSGQVRRLPDITVR"  
 4837. .5721  
 /locus\_tag="pp1326"  
 4837. .5721  
 /locus\_tag="pp1326"  
 /note="identified by match to PFAM protein family HMM  
 PF00590"  
 /codon\_start=1  
 /transl\_table=11  
 /product="tetracycline methylase family protein"  
 /protein\_id="AAN66950.1"  
 /db\_xref="GI:24982784"  
 /translation="MCAVTDVAGASKSTVGLYVATPIGNLDDMSARALKVLADVAL  
 TAAEDTRHSIRLLQHFQIDTPLAACHEHNERDEGRFLSKLLAGENVALVSDAGTPLI  
 SDPGTHLVQRQAAGSVVYPVPGACALIAALSAAGLPSDRIFEGFLPAKQARRARL  
 EQVREEPRTLIFYEAPHRILECLEDEMEAVFGDRLPALLARELTFTFELGLPLAELR  
 AFVAGDSNQRCQECVVLVGVWSAPEGEQAISAEARVLDLLALLAEPLKRAALAAEIT  
 GVRKNLLYQLALEKQKAB"  
 complement(5744. .6164)  
 /locus\_tag="pp1327"  
 /note="This region contains a gene with one or more  
 premature stops or frameshifts, and is not the result of a  
 sequencing artifact; conserved hypothetical protein,  
 degenerate; similar to GB:X16944, and SP:P15488;  
 identified by sequence similarity; putative"  
 6455. .6928  
 /locus\_tag="pp1328"  
 6455. .6928  
 /locus\_tag="pp1328"  
 /note="identified by match to PFAM protein family HMM  
 PF02381"  
 /codon\_start=1  
 /transl\_table=11  
 /product="conserved hypothetical protein TIGR00242"  
 /protein\_id="AAN66951.1"  
 /db\_xref="GI:24982786"  
 /translation="MWGITVFRGANAVSLDAKGRKLAMPSTRYRDELDSRCNGQLIVTI  
 DAVDPCLCVPLDEWEQTEAKRLALPSURENRRQLLIGNVLELDGSGRFLVPP  
 RLREYAKLDKKAMLVGQLNKFQLEDENAVSAADLAAIQPGAMPDELRLIL"  
 6925. .7872  
 /locus\_tag="pp1329"  
 6925. .7872  
 /locus\_tag="pp1329"  
 /note="similar to GP:9950652; identified by sequence  
 similarity; putative"  
 /codon\_start=1  
 /transl\_table=11  
 /product="conserved hypothetical protein TIGR00006"  
 /protein\_id="AAN66952.1"  
 /db\_xref="GI:24982787"  
 /translation="MTIDSGFNHTVLLDEAVEALALRADCYLDGTGFRGSHSLIL  
 SKLPQRLGLGDPKQAIATQALAAEDGRFVQVQSFAELGAEVARGHLKGVSGV  
 LLDGVSSPQDDDPERGFSLNDGPLDMNRRLQVSAAEIATAPVEETIARHVKYGV  
 EERFAGMARAVVERREKQPTRTADLAELVKVANPAWEKGNKATPAFCQGLRTHVNN  
 ELGDLEAGLAALDALEVGGRLAYISFHSLELDRIYKLFMRKLVKGEADNLPRLNPVQH  
 KVFEPKIKLIGKAQFAEAEELKANPRSRSAVMVAEKL"  
 7869. .8162  
 /gene="ftsL"  
 /locus\_tag="pp1330"  
 7869. .8162  
 /gene="ftsL"  
 /locus\_tag="pp1330"  
 /note="similar to GB:M11186, GB:M62611, SP:P01178,  
 PID:L89411, PID:386991, and PID:825699; identified by  
 sequence similarity; putative"

## Alignment Scores:

```

Pred. No.:      1.11e-108      Length:      301995
Score:          1835.00      Matches:      362
Percent Similarity: 88.62%      Conservatives: 35
Best Local Similarity: 80.80%      Mismatches: 51
Query Match:      81.19%      Indels:      0
DB:              1              Gaps:      0

US-09-701-229-2 (1-448) x AE016779 (1-301995)

Qy 1 MetSerLeuLeuAlaSerAspHisPheArgIleValValGlyLeuGlyLysSerGlyMet 20
Db ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
13843 GTGCTTGTATCGCTTCCGACCAATTCGCGATCGTTGTCGGCCCTCGCAAGAGCGCGATG 13902
Qy 21 SerLeuValArgTyrLeuAlaArgArgGlyLeuProPheAlaValValAspTyrArgGlu 40
Db ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
13903 TCCCTGGTTCCTTCCTGGCGAGCGCGGGGATTCGCTTTCGGCTCGCGATACCGCGAG 13962
Qy 41 AsnProGluLeuAlaThrLeuArgAlaGlnTyrProGlnValGluValArgCysGly 60
Db ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
13963 CAGCGCGCGAATCGGAACCCCTCGCGCTGATTACCGCGAGGTGGAAGTCGCGTGTGGT 14022
Qy 61 GluLeuAspAlaGluPheLeuCysSerAlaArgGluLeuTyrValSerProGlyLeuSer 80
Db ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
14023 GAATGACGTGGACTTCCTGTGCGCGCGCAACGAGCTGTACGTGAGCCCGCGCTGGCC 14082
Qy 81 LeuArgThrProAlaLeuValGlnAlaAlaLysGlyValArgIleSerGlyAspIle 100
Db ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
14083 TTGGCCACGGCGGCTTGCAGAGCGTCCGCGGTGGCGTGAAGCTGTGCGGACATA 14142
Qy 101 AspLeuPheAlaArgGluAlaLysAlaProIleValAlaIleThrGlySerAsnAlaLys 120
Db ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
14143 GAGCTGTTCCTGCCATGCCAAGCGCGGATGTTGCCATCAGTGTTCACAGCCCAAG 14202
Qy 121 SerThrValThrLeuValGlyGluMetAlaValAlaAlaAspLysArgValAlaVal 140
Db ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
14203 AGCACCTCAGCAGCGTGGTGGTGAATGGCGCCCAAGCGCGCAACGCGTGGCGGTG 14262
Qy 141 GlyGlyAsnLeuGlyThrProAlaLeuAspLeuAlaAspAspIleGluLeuTyrVal 160
Db ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
14263 GCGGTAACTCGTACCCCGGACCTGCTGCTGCGCGATGACATCGAGCTGTAGCTG 14322
Qy 161 LeuGluLeuSerSerPheGlnLeuGluThrCysAspArgLeuAsnAlaGluValAlaThr 180
Db ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
14323 CTGGAGCTTTCAGCTTCCAGCTGGAACCACTGACCACTCAACGCGGAAGTGGCGACC 14382
Qy 181 ValLeuAsnValSerGluAspHisMetAspArgTyrAspGlyMetAlaAspTyrHisLeu 200
Db ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
14383 GTTCTGAACATCAGCGAAGACCAATGAGCGCTATAGCGCGCTTCGCGCTACCACTG 14442
Qy 201 AlaLysHisArgIlePheArgGlyAlaArgGlnValValValAsnArgAlaAspAlaLeu 220
Db ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
14443 GCCAACAACCGGATTTCCCGGTGGCGCGGAGTGTGTGTCAACCGCAGAGCGCGCTG 14502
Qy 221 ThrArgProLeuLeuAlaAspThrValProCysTyrPheGlyLeuAsnLysProAsp 240
Db ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
14503 AGCGCGCGCTGCGGTGGAGCGCGCTCCATGCTGGACCTTCGSCCTCAACCGCGCGAC 14562
Qy 241 PheLysAlaPheGlyLeuIleGluAlaAspGlyGlnLysTyrPheAlaPheGlnPheAsp 260
Db ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
14563 TTCAAGACGTTTGGCTTCGCTGAAGTGAAGCGGTGAGAAATACCTGCGGTTCGAGTTCAG 14622
Qy 261 LysLeuLeuProValGlyLeuLysIleArgGlyAlaHisAsnTyrSerAsnAlaLeu 280
Db ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
14623 ACGTGTATGCGCGCGCTGAACTGAAGTTTCGGGTGCTCACAACCAAGAGTACGCGCTG 14682
Qy 281 AlaAlaLeuAlaLeuGlyHisAlaValAlaGlyLeuProPheAspAlaMetLeuGlyAlaLeu 300
Db ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
14683 GCGCGCTGCGCTGGGCGATGCTGCGCGCTGCGGTTCGAGCCCATGCTGGAGGCTCTG 14742
Qy 301 LysAlaPheSerGlyLeuAlaHisArgCysGlnTrpValArgGluArgGlnGlnValSer 320
Db ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::

```

```

Db 14743 CGCGAATTCGGTGGCTGGCCCGCCACCGGTGCCAGCGGTACGCGAGCGTAATGCGGTGAAC 14802
Qy 321 TyrTyrAspAspSerLysAlaThrAsnValGlyAlaAlaLeuAlaIleGluGlyLeu 340
Db ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
14803 TGTACGACGATTCCTCAAGACCAACAGCTGGTGGCAGCCCTGGCGGCATCGAAGGCTG 14862
Qy 341 GlyAlaAspIleAspGlyLysLeuValLeuAlaGlyLeuAlaGlyAspGlyLysGlyAlaAsp 360
Db ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
14863 GCGCGTGAATCGAAGGCAAGTTGGTGTCTGCGCGGTGGTGGCAGGCAAGGTCGTGAA 14922
Qy 361 PheHisAspLeuArgGluProValAlaAlaArgPheCysArgAlaValValLeuLeuGlyArg 380
Db ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
14923 TTCACGCTGCTGGCGGAGCGGTCACAGCGTTTCGCGCGCGGTGATTGCTTGGCGGT 14982
Qy 381 AspAlaGlyLeuIleAlaGlnAlaLeuGlyAsnAlaValProLeuValArgValAlaThr 400
Db ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
14983 GATGCCGAGCGCTGGCGGAGCGCTTGGCGGATGGCTGGCTGGCTGGCTGGCTGGCTGG 15042
Qy 401 LeuAspGluAlaValArgGlnAlaAlaGluLeuAlaArgGluGlyAspAlaValLeuLeu 420
Db ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
15043 CTTGACGATCGCTACAGCAGTGTGCGAACTGGCCAGCGCGGATGCGGTGCTGCTG 15102
Qy 421 SerProAlaCysAlaSerLeuAspMetPheLysAsnPheGluGluArgGlyArgLeuPhe 440
Db ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
15103 TCGCCAGCATGCGCGACGCTCGACATGTTCAAGAACTTCGAGGAACGGGACGCGCTGTT 15162
Qy 441 AlalysAlaValGluLeuAla 448
Db ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
15163 GCCGAGCAGCGGAGGCTTGCA 15186

RESULT 4
AE016871/c 311600 bp DNA linear BCT 05-MAR-2003
LOCUS Pseudomonas syringae pv. tomato str. DC3000 section 16 of 21 of the
DEFINITION complete genome.
ACCESSION AE016871 AE016853
VERSION AE016871.1 GI:28854552
SOURCE Pseudomonas syringae pv. tomato str. DC3000
ORGANISM Pseudomonas syringae pv. tomato str. DC3000
REFERENCE 1 (bases 1 to 311600)
AUTHORS Buell, R., Joardar, V., Khouri, H., Fedorova, N., Tran, B., Russell, D.,
Berry, K., Uterback, T., Van Aken, S., Feldblyum, T., Gwinn, M.,
Dodson, R., DeBoy, R., Durkin, A., Kolonay, J., Madupu, R.,
Daugherty, S., Brinkac, L., Beanan, M., Haft, D., Selengut, J.,
Neilson, W., Davidson, T., White, O., Fraser, C. and Collmer, A.
Complete Sequence of Pseudomonas syringae
Unpublished
TITLE 2 (bases 1 to 311600)
JOURNAL
AUTHORS Buell, R., Joardar, V., Khouri, H., Fedorova, N., Tran, B., Russell, D.,
Berry, K., Uterback, T., Van Aken, S., Feldblyum, T., Gwinn, M.,
Dodson, R., DeBoy, R., Durkin, A., Kolonay, J., Madupu, R.,
Daugherty, S., Brinkac, L., Beanan, M., Haft, D., Selengut, J.,
Neilson, W., Davidson, T., White, O., Fraser, C. and Collmer, A.
Direct Submission
TITLE Submitted (03-MAR-2003) The Institute for Genomic Research, 9712
JOURNAL Medical Center Dr, Rockville, MD 20850, USA
FEATURES
source
location/Qualifiers
1. 311600
/organism="Pseudomonas syringae pv. tomato str. DC3000"
/mol_type="genomic DNA"
/strain="DC3000"
/db_xref="taxon:223283"
/note="pathovar: tomato"
131..601
/gene="bfr"
131..601
/locus_tag="PSPT04160"
131..601
/gene="bfr"
131..601
/locus_tag="PSPT04160"
/note="similar to SP:P22759, and SP:P22759; identified by

```

[illegible]

1





```

/locus_tag="CBU0008"
/notes="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AA089578.1"
/db_xref="GI:29540634"
/translation="MSNFQKQKRAVSDSPALSNFFSIVFKSVTSIPVLSNWRRTN
PPKPPKLKLTQGRATFVS"
9560..10993
/gene="dacB"
/locus_tag="CBU0009"
9560..10993
/gene="dacB"
/locus_tag="CBU0009"
/notes="identified by match to PFAM protein family HMM
PF00905"
/codon_start=1
/transl_table=11
/product="D-alanyl-D-alanine
carboxypeptidase/D-alanyl-D-alanine-
endopeptidase"
/protein_id="AA089579.1"
/db_xref="GI:29540635"
/translation="MKPPYLAILLSCPLFAVEKQNPQTNILILSIKSNILG
FIVSDKTRGVIFERSQELFSPASTKULFVAVALYILGSDYQFTALLTNGAIRGQ
TLQGNLTLPFGSDPELTEDLNRLIEKLKELGTHRSGHVYIDNTAYNDVPPGWML

Alignment Scores:
Pred. No.: 1 58e-55 Length: 301727
Score: 1023.00 Matches: 210
Percent Similarity: 64.47% Conservativeness: 84
Best Local Similarity: 46.05% Mismatches: 140
Query Match: 45.27% Indels: 22
DB: 1 Gaps: 3
US-09-701-229-2 (1-448) x AE016960 (1-301727)

```

```

Qy 2 SerLeuLeaSerAspHisPheArgIleValValGlyLeuGlyLysSerGlyMetSer 21
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
Db 118030 ACGCTTATGATGACCAATTAACGGTTATCGTTGGTTAGTAAACGGGATATCT 118089
Qy 22 LeuValArgTyrLeuAlaArgGlyLeuProPheAlaValValAspThrArgGluAsn 41
|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
Db 118090 TGTGCGGAGTTTTAGCGAAGAACACCGCGTTGCGGTGATGGACATCTGTGAAGAA 118149
Qy 42 ProProGluLeuAlaThrLeuArgAlaGlnTyrProGlnValGluValArgCysGlyGlu 61
|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
Db 118150 CCCCCGAGTGGGAAATTTATCAAACTTACCTCGCTCGAATGATCCGCGGCCAA 118209
Qy 62 LeuAspAlaGluPheLeuCysSerAlaArgGluLeuTyrValSerProGlyLeuSerLeu 81
::: ||| |||:::|||||:::|||||:::|||||:::|||||:::|||||
Db 118210 TTTTCGGAATAATTATGAACGAAGCGCAAGAGATTTTAAAGTCTGCGGTTGCTG 118269
Qy 82 ArgThrProAlaLeuValGlnAlaAlaLysGlyValArgIleSerGlyAspIleAsp 101
::: ||| |||:::|||||:::|||||:::|||||:::|||||:::|||||
Db 118270 CAGAGCGGTAAATGCGCAAGCAAGCGCGCAAGCAAAATTCATTTATGCGGATATGAA 118329
Qy 102 LeuPheAlaArgGluAlaLysAlaProIleValAlaIleThrGlySerAsnAlaLysSer 121
|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
Db 118330 TTATTCGACGCAATGTGATAAACCAATTATGCTATTACCGCTCCATGCGCAAAACC 118389
Qy 122 ThrValThrThrLeuValGlyGluMetAlaValAlaAspLysArgValAlaValGly 141
|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
Db 118390 ACCGTGACGCGGTAGTGGGATTAATGATGAAGCGCGCGCAACAGTTTCAGTTTCG 118449
Qy 142 GlyAsnLeuGlyThrProAlaLeuAspLeuLeuAlaAspIleGluLeuTyrValLeu 161
|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
Db 118450 GGTAATATGGCGCGCGGTCTTAGACAAATACACCTGACCGCGGATTTATGCTACTG 118509
Qy 162 GluLeuSerSerPheGlnLeuGluThrCysAspArgLeuAsnAlaGluValAlaThrVal 181
|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
Db 118510 GAAATATCGAGCTTCCAACTTGAAACGACATTTTCTTTACGATCTCAAGCGGCAACATA 118569

```

```

Qy 182 LeuAsnValSerGluAspHisMetAspArgTyrAspGlyMetAlaAspTyrHisLeuAla 201
|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
Db 118570 TTAATATTAGCGAAGACCATATGAATCGTATCGGACTTACAAGATTATCTCCGAGCT 118629
Qy 202 LysHisArgIlePheArgGlyAlaArgGlnValValValAsnArgAlaAspAlaLeuThr 221
||| |||:::|||||:::|||||:::|||||:::|||||:::|||||
Db 118630 AAACAACGGCATTTATCTAGTTGTTTATTCCCATTCGTAAC----- 118671
Qy 222 ArgProLeuIleAlaAspThrValProCysTrp----- 232
Db 118672 -----CGGACGAAACAGAAATTTTCGGCCCATCTTCGGTTTATAAAAAACCT 118719
Qy 233 ---SerPheGlyLeuAsnLysProAspPheLysAlaPheGlyLeuIleGluAspGly 251
|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
Db 118720 TTGCTTTTCGGGTTAAAT-----AATGCTGCTGATTTTCGTGGCAGAACATAATCAA 118773
Qy 252 GlnLysTrpLeuAlaPheGlnPheAspLysLeuProValGlyGlnLeuLysIleArg 271
::: |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
Db 118774 AAAACGCTATTTCCTATCATCGGTAAGATTTTATGCCCATCCAAAGAAATTTAAATTAAC 118833
Qy 272 GlyAlaHisAsnTyrSerAsnAlaLeuAlaLeuAlaLeuGlyHisAlaValGlyLeu 291
|||::: |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
Db 118834 GCCCGTCATCATCTCCAAATGACACTTGTGCTATGACATTTGGACAGCAGCAGTAAATA 118893
Qy 292 ProPheAspAlaMetLeuGlyAlaLeuLysAlaPheSerGlyLeuAlaHisArgCysGln 311
||| ::| ||||| |||::: |||||::: |||||::: |||||::: |||||::: |||||
Db 118894 CCCATAGAAAATATGCTTTCATGATTACGTGATTTCGGTATTCGCCACCGCTGTCAA 118953
Qy 312 TrpValArgGluArgGlnGlyValSerTyrTyrAspSerLysAlaThrAsnValGly 331
|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
Db 118954 TGGGTACGCAAGTATAAGGAAATGTGATTATTAATGATTCAAAAGGCACAAACGTTGGC 119013
Qy 332 AlaAlaLeuAlaAlaIleGlyLeuGlyAlaAspIleAspGlyLysLeuValLeuLeu 351
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 119014 GCTACCGCGCTGCCATTGAAGCTTAGGCGCAAGGCAAAAGGCAACTTAATCCTATC 119073
Qy 352 AlaGlyCysAspGlyLysGlyAlaAspPheHisAspLeuArgGluProValAlaArgPhe 371
|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
Db 119074 GCTGGGGACAAGCAGGCGCGGATTTTCTCATTAAGAACGCTGCTAAACGCTTAC 119133
Qy 372 CysArgAlaValValLeuLeuGlyArgAspAlaGlyLeuIleAlaGlnAlaLeuGlyAsn 391
::: |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
Db 119134 GTTAACAAAGTATTTAATCGGAGGATGCTCCGCTGTGTAAGAAACCTTAAGAA 119193
Qy 392 AlaValProLeuValArgValAlaThrLeuAspGluAlaValArgGlnAlaAlaGluLeu 411
::: |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
Db 119194 ATAACGCTTATTAACACGCTGATTCAATGACGAAACCGCTCAACGATCCATCAAGCC 119253
Qy 412 AlaArgGluGlyAspAlaValLeuLeuSerProAlaCysAlaSerLeuAspMetPheLys 431
|||::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 119254 GCCAAAGCCGCGATATGTTTATTATCCCGCTTGCAGGATTTTGATATGTTTACC 119313
Qy 432 AspPheGluGluArgGlyArgLeuPheAlaLysAlaValGluLeu 447
|||||:::||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 119314 AACTAGAACACCGCGCGATGTTTTTACAGAAACGCTGGAAGCGCTT 119361

```

## RESULT 6

AL627265

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

AL627265 251050 bp DNA linear BCT 06-JUN-2002  
 Salmonella enterica serovar Typhi (Salmonella typhi) strain Cr18,  
 complete chromosome, segment 1/20.  
 AL627265 AL513382  
 AL627265.1 GI:16501283

Salmonella enterica subsp. enterica serovar Typhi  
 Salmonella enterica subsp. enterica serovar Typhi  
 Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 Enterobacteriaceae; Salmonella.  
 1 (bases 1 to 251050)  
 Parkhill, J., Dougan, G., James, K. D., Thomson, N. R., Pickard, D.,  
 Wain, J., Churcher, C., Mungall, K. L., Bentley, S. D., Holden, M. T. G.,  
 Sebahia, M., Baker, S., Basham, D., Brooks, K., Chillingworth, T.,  
 Connerton, P., Cronin, A., Davis, P., Davies, R. M., Dowd, L., White, N.,

Farrar,J., Feltwell,T., Hamlin,N., Haque,A., Hien,T.T., Holroyd,S., Jagels,K., Krogh,A., Larsen,T.S., Leather,S., Moule,S., O'Gaora,P., Parry,C., Quail,M., Rutherford,K., Simmonds,M., Skelton,J., Stevens,K., Whitehead,S. and Barrell,B.G.  
Complete genome sequence of a multiple drug resistant Salmonella enterica serovar Typhi CT18

Nature 413 (6858), 848-852 (2001)

21534947

2 (bases 1 to 251050)

11677608

Parkhill,J.

Direct Submission

Submitted (25-Oct-2001) Submitted on behalf of the Salmonella

sequencing team, Sanger Centre, Wellcome Trust Genome Campus,

Hinxton, Cambridge CB10 1SA, UK

E-mail: parkhill@sanger.ac.uk

Notes:

Details of S. typhi sequencing at the Sanger Centre are available

on the World Wide Web.

(URL, [http://www.sanger.ac.uk/Projects/S\\_typ1/](http://www.sanger.ac.uk/Projects/S_typ1/)).

Location/Qualifiers

1. 251050

/organism="Salmonella enterica subsp. enterica serovar

Typhi"

/mol\_type="genomic DNA"

/strain="CT18"

/db\_xref="taxon:90370"

190. 255

/gene="STY0001"

/note="synonym: thrL"

190. 255

/gene="STY0001"

/note="Orthologue of E. coli thrL (LPT\_ECOLI); Fasta hit

to LPT\_ECOLI (21 aa), 86% identity in 21 aa overlap"

/codon\_start=1

/transl\_table=11

/product="thr operon leader peptide"

/protein\_id="CAD01154.1"

/db\_xref="GI:16501284"

/db\_xref="SPTREMBL:Q8XG12"

/translation="MNRISTTITITITITGGAG"

337. 2799

/gene="STY0002"

/note="synonym: thrA"

337. 2799

/gene="STY0002"

/note="Orthologue of E. coli thrA (AKIH\_ECOLI); Fasta hit

to AKIH\_ECOLI (820 aa), 94% identity in 820 aa overlap"

/codon\_start=1

/transl\_table=11

/product="aspartokinase I/homoserine dehydrogenase I"

/protein\_id="CAD01155.1"

/db\_xref="GI:16501285"

/db\_xref="SPTREMBL:Q8Z9T7"

/translation="NRVLKFGSTVANAERELRVADILLESROGQVATVLSAPAKIT

NHLVAMIEKTIQQDALPNISDAERIFSDLLAGLSAQGGPFLARKMVKQVEQAQIK

LYHRSILSGQCPDSINALICRGKMSIATMAGLEAGRHVTVIDPVELAVGHI

LESTVDIAESTRIASQIPADHMTLMAGFTAGNKGELVLRGRNSDYSAVLAAQY

RADCCIEIWTVDVGYTCDPQVPDARLLKMSYQEAELSIFGAKVLHPRTITPAQF

QTPCLIKNTGNPQAPTGLIGASDDNLPVKGISNLNNMFSVSGPGMGIMGAAR

VFAAMSRAISVLLITQSSSEYISFCVPSQDCARARAMQDEEYLEKEGLLEPLAV

TERLAISVVGDMTTLRGISAKFPAALARANINIVATQSSERSISVYVNNDDATT

GVRVTHQMLFNDQVIEVYIGVGVGGALLEQURQQTWLKNKHIDURVCGVANSKA

LLTNVHGLNMQWQALQAAPFNGLRIILVKEYHLNPNIVDCTSSQAVAOYAD

FLREGFVYVTPNKNANTSMFYHQLRFAAQSRKFLYDNTVGAGLPVIEINQLLN

AGDELFQSGILSSLSIFGKLEGMSLSQATALAREMYTEPDPRDLGMDVARK

LLILARETGLSDSDIVIEPVLDPEDFASGDVTFAMHLPLQDDAFARVAKARDEC

KVLRYVGNIEEDGVCVKIAEVDGNDPLFKVNGENALAFYSHYQPLPLVLRGYGAG

NDVTAAGVFADRLRLTSLWKLWG"

337. 1101

/gene="STY0002"

/note="Pfam match to entry PF00696 aakinese, Amino acid

kinase family, score 207.70, E-value 1.7e-58"

misc\_feature

misc\_feature

misc\_feature

gene

CDS

misc\_feature

misc\_feature

gene

CDS

misc\_feature

misc\_feature

gene

CDS

343. 369  
/gene="STY0002"  
/note="PS00324 Aspartokinase signature"  
1819. 2781  
/gene="STY0002"  
/note="Pfam match to entry PF00742 Homoserine\_dh,  
Homoserine dehydrogenase, score 535.60, E-value 3.4e-157"  
2314. 2382  
/gene="STY0002"  
/note="PS01042 Homoserine dehydrogenase signature"  
2801. 3730  
/gene="STY0003"  
/note="synonym: thrB"  
2801. 3730  
/gene="STY0003"  
/note="Orthologue of E. coli thrB (KHSE\_ECOLI); Fasta hit  
to KHSE\_ECOLI (310 aa); 94% identity in 308 aa overlap"  
/codon\_start=1  
/transl\_table=11  
/product="homoserine kinase"  
/protein\_id="CAD01156.1"  
/db\_xref="GI:16501286"  
/db\_xref="SWISS-PROT:Q8XCP5"  
/translation="MKVYAPASSANMSVGFVGLAAVTPVDGTLGLDVSVEAADHF  
RLNLGREFADKLPPRENIYQWRECFQALGKTIPTVMLEKNMPTGSLGSSACS  
VVAALVAMNEHCGLPNDRLIALMGEIRISGSIHYDNDVAPCFGLGMQIMIEGNI  
ISQOVPGFDEMLWVLAIPGIVKIVSTAEARAILPAQVRRDCTAHGRLAGFTACYSRQ  
POLAALMKDVIAEPYRARIPLPGFSQARQAVSEIGALASIGSGSTFLFALCDKPEIA  
QRVADWLSKHVQLNQEGFVHCRLDTAGRVVG"  
3005. 3196  
/gene="STY0003"  
/note="Pfam match to entry PF00288 GHMP\_Kinases, GHMP  
kinases putative ATP-binding proteins, score 79.20,  
E-value 5.7e-21"  
3068. 3103  
/gene="STY0003"  
/note="PS00627 GHMP kinases putative ATP-binding domain"  
3734. 5020  
/gene="STY0004"  
/note="synonym: thrC"  
3734. 5020  
/gene="STY0004"  
/note="Orthologue of E. coli thrC (THRC\_ECOLI); Fasta hit  
to THRC\_ECOLI (428 aa), 93% identity in 428 aa overlap"  
/codon\_start=1  
/transl\_table=11  
/product="threonine synthase"  
/protein\_id="CAD01157.1"  
/db\_xref="GI:16501287"  
/db\_xref="SPTREMBL:Q8Z9R6"  
/translation="MKLYNLKDHNEQVSFAQVTOGLGKQOGLFFPHDLPFSLTEID  
EMLNQDFVRSKILISAFIDETIPQILIEERVRAAFAPAPVAQVDSVGLGELFHGP  
TLAFKDFGGRFMAQMLTHISGDKPVTILTSGDTGAVAHAFYGLENNRVVILYPRG  
KISLPQEKLFCTLGSIETVAIDGDFDCAQALVKQADDEELKTAGLNSANSINIS  
LLAQICFIEVAQIPLQCARNLQIVISVPNGFGLTAGLLAKSLGLPKVRIATATNIN  
DTVPRFLHDGKWAQATQATLSNADWVSQPNWPRVEELFRKKIWRULFELGYAAVDDT  
TQTQTMRELKAGYIISEPHAAYRALRDNGEYGLFGLGTAHPAKFKESVESILGE  
TLALPEALAEARADLPILLSHLPADFAALRKLMTRQ"  
3962. 4867  
/gene="STY0004"  
/note="Pfam match to entry PF00291 PALP,  
Pyridoxal-phosphate dependent enzyme, score 104.50,  
E-value 2e-27"  
4022. 4066  
/gene="STY0004"  
/note="PS00165 Serine/threonine dehydratases  
pyridoxal-phosphate attachment site"  
complement(5114. 5887)  
/gene="STY0005"  
/note="synonym: yaaA"  
complement(5114. 5887)  
/gene="STY0005"  
/note="Orthologue of E. coli yaaA (YAAA\_ECOLI); Fasta hit

to YAAA\_ECOLI (258 aa), 86% identity in 257 aa overlap"

```
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="CAD01158.1"
/db_xref="GI:16501288"
/db_xref="SPRMBL:Q8Z9R5"
/transl_table="MULISPAKTLDYQSLATRYTQPELLDHSQQILQOARQLSP
QISRLMISKLADLNATRFDMQPHFTPDNARQAILAFKGDVYTGLOAETFDNADF
FAOHLKMLSLGVLRLEMLQYRLEMLPRGKDYOFWGDITDKNEALE
AQGRVVNVLASREYFKSVKPKLDALIKPFLDEKNGFKVSVFYAKKARGLMSRF
LIENRLKPOLFAFDREGYFDEESTQDELVEKRYEQ"
complement(5966..7396)
/gene="STY0006"
/note="synonym: yaaJ"
```

gene

CDS complement(5966..7396)

```
/gene="STY0006"
/note="Similar to Bacillus subtilis amino acid carrier
protein alst ALST SW:ALST_BACSU (Q45068; P40743) fasta
scores: E(): 0, 45.3% id in 477 aa, and to Alteromonas
haloplanktis na(+)-linked D-alanine glycine permease DAGA
SW:DAGA_ALTHA (P30144) fasta scores: E(): 0, 36.5% id in
315 aa
```

Orthologue of E. coli yaaJ (YAAJ\_ECOLI); Fasta hit to  
YAAJ\_ECOLI (476 aa), 76% identity in 476 aa overlap"

```
/codon_start=1
/transl_table=11
/product="putative amino-acid transport protein"
/protein_id="CAD01159.1"
/db_xref="GI:16501289"
/db_xref="SPRMBL:Q8Z9R4"
```

```
/transl_table="MPEFFSFINELMGSMVYLLGAGCWFTWRTGFIQFVIRQFS
RSKLSLSPQGGTSLQACTSLAARGSGNLGVALAIAAGPGGAVFWMVSAIIG
MATSFACSLAQLYKQDPDQFGRGPAWYMGRLGVMWVFPALPLVAYGLIFS
VOANVSRMLAFNAPLISGIALAFCALLIIRGIKGVARLWMLPIPLIALWVAG
SLQALWHIQMGPVIASIVKSAFGWQERAGAGYTLQAITSGFORGESEAGAG
STPNAAAATSPHPVQAQIVQIMGVFSDTIICHTASAMILLIAGNASHSSTEGIQ
LLQHWVSLTGEWGASEVALIVILFSSIVANITYAENNLFLRLHNAKAIWLLRIA
TLGHVIASTLISPLIWLQADIMACMAITNLTAILLLSVPVVTLAGDYLRLKLGVR
```

#### Alignment Scores:

Pred. No.:	1.83e-53	Length:	251050
Score:	990.00	Matches:	233
Percent Similarity:	64.18%	Conservative:	59
Best Local Similarity:	51.21%	Mismatches:	139
Query Match:	43.81%	Indels:	24
DB:	1	Gaps:	10

US-09-701-229-2 (1-448) x AL627265 (1-251050)

Qy	2	SerLeuIleAlaSerAspHisPheArg-----lIleValValGlyLeuGlyLysSer	18
Db	148805	ACGTTAATCATGCTGATTACACAGGACAAAACGTCGTCATATATCGTCTGGGCTTAACC	148864
Qy	19	GlyMetSerLeuValArgTyrLeuAlaArgArgGlyLeuProPheAlaValAlaValAspThr	38
Db	148865	GGACTCTCTTCGCTGGAGCTTTTCTCGCCGCGCGGTGACGCGCGGTGATGGATACT	148924
Qy	39	ArgGluAsnProProGluLeuAlaThrLeuAlaGlnTyrProGln---ValGluVal	57
Db	148925	CGCTGACGCGCGCGGTCTGGATAAGCTG-----CCGCAAGAGGTGTAGCGT	148972
Qy	58	ArgCysGlyGluLeuAspAlaGluPheLeuCysSerAlaArgGluLeuTyrValSerPro	77
Db	148973	CACGTTGGCGCTGACAGCAGTGGCTCTTACGCGGGAATTAATCGTCGCCAGCCCT	149032
Qy	78	GlyLeuSerLeuArgThrProAlaLeuValGlnAlaAlaLysGlyValArgIleSer	97
Db	149033	GGTATTGCTCGCGCATCCCTCTCTTAGCGCTCCGCCAGCGCTGGAGTGGAAATGTC	149092
Qy	98	GlyAspIleAspLeuPheAlaArgGluAlaLysAlaProIleValAlaIleThrGlySer	117
Db	149093	GGCGATATCGAACTGTTTGGCCGGAAGCGCAGCGCGGATTTGTGGCCATCACCGGCTCG	149152

Qy	118	AsnAlaLysSerThrValThrThrLeuValGlyGluMetAlaValAlaAlaAspLysArg	137
Db	149153	AACGGCAAAAGCACCGTGACCACTTAGTGGCGAGATGGGAAGCGCGGCGGCTCAAT	149212
Qy	138	ValAlaValGlyGlyAsnLeuGlyThrProAlaLeuAspLeuLeuAlaAspIleGlu	157
Db	149213	GTCGGCGTTGGCGCAATATCGGCTTCGCGCGCTGATGCTGTGGATGCGGATCGCGAA	149272
Qy	158	LeuTyrValLeuGluLeuSerSerPheGlnLeuGluThrCysAspArgLeuAsnAlaGlu	177
Db	149273	TTGTACGTGCTGGAAATATCCAGTTTCCAATTTGAAACTACCTCAAGTTTGCACGCGCG	149332
Qy	178	ValAlaThrValLeuAsnValSerGluAspHisMetAspArgTyrAsp---GlyMetAla	196
Db	149333	CGCGCAACCGTGCTCAACGTCACCTGAAGATCATATGACCGGTATCCGTTTGTGTTGCA	149392
Qy	197	AspTyrHisLeuAlaLysHisArgIlePheArgGlyAlaArgGlnValValAlaAsnArg	216
Db	149393	CAGTACCGCGCGCGCAACTGCGCTCTACGAGAAGCGCAAGTGTGCTAGTAGTGAATGCC	149452
Qy	217	AlaAspAlaLeuThrArgProLeuIleAlaAspThrValProCysTyrSerPheGlyLeu	236
Db	149453	GATGACCGCTTGACGATGCGGTACGCGGGGCCCATGAGCGCTCGCTGAGCTTTGGCGTC	149512
Qy	237	AsnLysProAspPheLysAlaPheGlyLeuGluAspGlyGlnLysTyrPheAla	256
Db	149513	AATATGGGTGATTATCAC-----CTTAATCGTCAGCAGGCGCAACCTGCTACGA	149563
Qy	257	PheGlnPheAspLysLeuLeuProValGlyGluLeuLysIleArgGlyAlaHisAsnTyr	276
Db	149564	GTCAAAGGTGAGAAGTGTCAACGTGAAGAGATGAAGCTTCGCTGACGATTAACCTAT	149623
Qy	277	SerAsnAlaLeuAlaLeuAlaLeuGlyHisAlaValGlyLeuProPheAspAlaMet	296
Db	149624	ACCATCGGTAGCGCGCTGCGCTGCGGATGCGGTAGCTGCTGCGCGCGCGCAGCAGT	149683
Qy	297	LeuGlyAlaLeuLysAlaPheSerGlyLeuAlaHisArgCysGlnTyrPvalArgGluArg	316
Db	149684	TTGAAGCGTTGACGACATTTACCGGTCTCGCGCATCGCTTCCAGCTGCGCTTGAGCAT	149743
Qy	317	GlnGlyValSerTyrTyrAspSerLysAlaThrAsnValGlyAlaAlaLeuAlaAla	336
Db	149744	AACGCGCTGCTGTCGATCAACGACTCGAAGACGACCAATGTCGCGAGTACGGAAGCGCG	149803
Qy	337	IleGluGlyLeuGlyAlaAspIleAspGlyLysLeuValLeuLeuAlaGlyGlyAspGly	356
Db	149804	CTAAACGCTTG-----CATGTGACGCTAGCTGCTGCTGCTGCTGCGCGCGCGCAGCT	149857
Qy	357	LysGlyAlaAspPheHisAspLeuArgGluProValAlaArgPheCysArgAla-----	374
Db	149858	AAGTCGCGCAGACTTT-----TCTCCGCTGGCGCTATCTGACCGCGCATCGT	149905
Qy	375	-----ValValLeuLeuGlyArgAspAlaGlyLeuIleAlaGlnAlaLeuGlyAsnAla	392
Db	149906	ATCCGCTGCTATTCCTTTGGCGCGGATGGCGCGCAGCTGTC---GCATCGCTCCGAA	149962
Qy	393	ValProLeuValArgValAlaThrLeuAspGluAlaValArgGlnAlaAlaGluLeuAla	412
Db	149963	ATC---GCCCAACAGACTGAGACGATGGAGAGCGGATGCTGCTGGCGCGCCGACGTT	150019
Qy	413	ArgGluGlyAspAlaValLeuSerProAlaCysAlaSerLeuAspMetPheLysAsn	432
Db	150020	CAGCGCGGTGATATGGTGTCTTCGCCGCCCTGCGCCAGCTCGATCATGCTTTAAAAAT	150079
Qy	433	PheGluGluArgGlyArgLeuPheAlaLysAlaValGluGluLeu	447
Db	150080	TTTGAGCAACGGGCGCATGCTTTTACCCTGCTGCGGAAGGAGTTA	150124

RESULT 7

AE016834

LOCUS

DEFINITION

Salmonella enterica subsp. enterica serovar Typhi Ty2, section 1 of

300169 bp

DNA

linear

BCT 21-MAR-2003

16 of the complete genome.  
AE016834 AE014613  
VERSION  
AE016834.1 GI:29136167  
SOURCE  
ORGANISM  
Salmonella enterica subsp. enterica serovar Typhi Ty2  
Salmonella enterica subsp. enterica serovar Typhi Ty2  
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
Enterobacteriaceae; Salmonella.  
Deng, W., Liou, S.-R., Plunkett III, G., Mayhew, G.F., Rose, D.J.,  
Burland, V., Kodoyianni, V., Schwartz, D.C. and Blattner, F.R.  
Comparative Genomics of Salmonella enterica Serovar Typhi Strains  
Ty2 and CT18  
J. Bacteriol. 185 (7), 2330-2337 (2003)  
MEDLINE  
22531367  
PUBMED  
12644504  
REFERENCE  
Deng, W., Liou, S.-R., Plunkett, G. III, Mayhew, G.F., Rose, D.J.,  
Burland, V., Kodoyianni, V., Schwartz, D.C. and Blattner, F.R.  
Direct Submission  
Submitted (25-SEP-2002) Laboratory of Genetics, University of  
Wisconsin - Madison, 445 Henry Mall, Madison, WI 53706, USA  
FEATURES  
source  
1. 300169  
/organism="Salmonella enterica subsp. enterica serovar  
Typhi Ty2"  
/mol\_type="genomic DNA"  
/strain="Ty2"  
/serovar="Typhi"  
/sub\_species="enterica"  
/db\_xref="taxon:209261"  
190..255  
/gene="thrL"  
/locus\_tag="t0001"  
190..255  
/gene="thrL"  
/locus\_tag="t0001"  
/note="corresponds to STY0001 from Accession AL513382:  
Salmonella typhi CT18"  
/codon\_start=1  
/transl\_table=11  
/product="thr operon leader peptide"  
/protein\_id="AAO67735.1"  
/db\_xref="GI:29136168"  
/translation="MNRISTTITITITITITGNGAG"  
337..2799  
/gene="thrA"  
/locus\_tag="t0002"  
337..2799  
/gene="thrA"  
/locus\_tag="t0002"  
/note="corresponds to STY0002 from Accession AL513382:  
Salmonella typhi CT18; homoserine dehydrogenase I"  
/codon\_start=1  
/transl\_table=11  
/product="aspartokinase I"  
/protein\_id="AAO67736.1"  
/db\_xref="GI:29136169"  
/translation="MRVLKFGTGVANAERFLRVADILESNRQGVATVLSAPAKIT  
NHLVAMIEKIGQDAPNISAERIFSDLLAGLASAQPGLPRLKRMVQEFAQIK  
HYLGISILGQCPDSINAALICRGEKMSIATMAGLEAGHRVTVIDPVEKILAVGHY  
LFSTVDIAIAISOIPADHMTLMAGFTAGNEKVELVILGRNGSDYSRAVLAACL  
RADCEIITWDVGYTCDPQVPDARLLKMSQDAMELSYFGARVLRPTITPAQF  
QIPCLIKTGNPQAPETLIGASSDDNLPVKGISNLNNMAMFVSQGMKGMIGMAAR  
VFAAMRAGISVLTITQSSEYSISFCVQSDCARARMQDEFFYLEKGLLEPLAV  
TERLAIISVVDGDMRTLRGISAKFFAALANINIVAIAQSSERSISVWVNDATT  
GVRTVHQLMFDVIEVYIGVGGVALLLEOLKQOITLKNKHIDIRVCGVANSKA  
LITNVHGLNDNQAELAAQAPNLRILRVKEVHLINPVIVDCTSSQAVADYAD  
FLREGFHVTPNKKANTSMDYHQLRFAAQRKFLIDYVNGAGLVPIENQLNLN  
AGDELQKSGILSGLSLFTFGKLEGMSLSQATALAREMGYTEPDRDLSGMDVARK  
LILARETGRELELSDIVIEVLPDEFDSGDVTFAMAHLPOLDODFAARVAKARDEG  
KYLTVGVNIEEDGVGRVIAEVDGNDPLFKVKNGENALAFYSHYYQPLPLVLRGYGAG  
NDVTAAGVFADLLRLTSNKLGV"

gene  
2801..3730  
/gene="thrB"  
/locus\_tag="t0003"  
2801..3730  
/gene="thrB"  
/locus\_tag="t0003"  
/note="corresponds to STY0003 from Accession AL513382:  
Salmonella typhi CT18"  
/codon\_start=1  
/transl\_table=11  
/product="homoserine kinase"  
/protein\_id="AAO67737.1"  
/db\_xref="GI:29136170"  
/translation="MVKYVAPASSANMSVGFVGLGAAVTPVDGTLIGDVVSVYAADHF  
RLHNLGRFADKLPPENINIVYQWERFQALGKTPVAMTLEKNMIGSGIGSSACS  
VVAALVAMNEHCGKPLNDPRLALLALMGELEGRISGSIHVDNVPACFLGNGQLMIENGI  
ISQVPGFDEWLWLYAPGKRTAEARAILPAQRQDCIAHGRHLAGFIHACISRO  
POLAALMKDVIJAEPIYRARLLPFGFQARQAVSEIGALASGISGSGPTLFCDCRKPETA  
QRVADWLKSHYLQNGEGFVHICRLDTAGARVVG"  
3734..5020  
/gene="thrC"  
/locus\_tag="t0004"  
3734..5020  
/gene="thrC"  
/locus\_tag="t0004"  
/note="corresponds to STY0004 from Accession AL513382:  
Salmonella typhi CT18"  
/codon\_start=1  
/transl\_table=11  
/product="threonine synthase"  
/protein\_id="AAO67738.1"  
/db\_xref="GI:29136171"  
/translation="MKLYNLKDHNEQVSFAQAVTQGLGKQGLFPPHDLPEPSLSTEID  
EMLNODFVSRSAKILSAFIDBEIPQOILEERVAAFAFAPVAQVESDVGCLELPHGP  
TLAPKDFGREGMAOMLTHISGDKPVTILTATSGDTGAFAVAHAFYGLNVVVILYPRG  
KISPLQKELCTLCGNIETVALDGFDAQALYKQAFDDEELKLTALGNSANSINIS  
LLAQICYFEAVAPQPGARNLQVIVSFGNFGLTAGLAKSLPLPVKRRFTAAATNIN  
DTQPTMRELKAGWATQATVSNAMDVSNQNNPVRVELEFRKIMRLTELGYAAVDDT  
TTLALPEALAERADLPLLSHLHPADFAALRLKLMTRQ"  
complement(5114..5887)  
/gene="yaaJ"  
/locus\_tag="t0005"  
complement(5114..5887)  
/gene="yaaJ"  
/locus\_tag="t0005"  
/note="corresponds to STY0005 from Accession AL513382:  
Salmonella typhi CT18"  
/codon\_start=1  
/transl\_table=11  
/product="conserved hypothetical protein"  
/protein\_id="AAO67739.1"  
/db\_xref="GI:29136172"  
/translation="MLILISPAKTLDYOSPLATTRYTOPELLDHSQQLIQAROLASAP  
QISRLMGISDKLADLNATRFHDWPHFTPDNARQAILAFKGDVYVTLGLOAFENDADF  
FAQHRLMSGLYGLVRDLDMQPYRLEMGINLENPRCKDLYQFGDITITDKLNPALE  
AQKRVVNVNLAASEEYFKSKPKKDLDAELIKPVFLDERKNGKRVVSVFYAKKARGLSRF  
ITENRLTKPEQLTAFDREGYFFDEETSTQDELVKRYEQ"  
complement(5966..7396)  
/gene="yaaJ"  
/locus\_tag="t0006"  
complement(5966..7396)  
/gene="yaaJ"  
/locus\_tag="t0006"  
/note="corresponds to STY0006 from Accession AL513382:  
Salmonella typhi CT18"  
/codon\_start=1  
/transl\_table=11  
/product="putative amino-acid transport protein"  
/protein\_id="AAO67740.1"  
/db\_xref="GI:29136173"  
/translation="MPEFFSFINEILWGSVMYILLGAGCWFTWRTGFIQFRIYTRQFS  
RSLKGSLSQPQGLTSFQALCTSLAARIGSGNLAGVALIAAGGPGVFWMWVSAIIG

MATSFACSLAQYIKERDPTGQFRGGPANYMARGLMRMWGVVFAFLVAYGLIFNS  
VQANASVRLAFNFIPIPLISGIALAFCALLIIRIGIKGVARLMOWLIPLIALIMVAG  
SVFTCLWHIEQMEGVIASIVKSAFQWQEAAGAAGYTLQOAITSGFORGFMSNEAGMG  
STNAAATSYPPHPVAQIGVOMIGVSDTIICITASAMILLAGNHASHSSTEGIQ  
LLOHAMVSLGEGASGVALLIVILFAFSIVANYIYAENNLFFLRHNKAKIWLRLA  
TLGVINGTILISPLIQLADNMACMAITNLTAILLSPVITLAGDYLQRKRLGVR  
PQDPRFPDIEPOLAPDWTDAASRD

gene

7665..8618

/gene="tal8"

/locus\_tag="t0007"

7665..8618

/gene="tal8"

/locus\_tag="t0007"

/note="corresponds to STY0007 from Accession AL513382"

Salmonella typhi CT18"

/codon\_start=1

/transl\_table=11

/product="transaldolase B"

/protein\_id="AA067741.1"

/db\_xref="GI:29136174"

/translation="MTDKLSLRQFTTVVADTGDIAAMKLYQPQDATTNPISLIINAAQ  
IPEYKLIIDDAVAKQOQSDRAQOVVDATDKLAVNIGLEILKLVPGRISTEVDARLS  
YDIEASTAKAKRIKLYNDAGISNDRILLKLASTWQGIKRAEQLERKGINCLTILES  
PAQARCAEAGVYLISFPVGRILDWYKANTDKDPAEDPGVSVTEIYEYKQHSY  
ETVVMGASFRNVEILELAGCDRLTIAPALLKEIAESEGAIERKLSFSGEVKARPERI  
TEAEFLWQHODPMVADKLADGIRKFAVDQEKLEKMGIDLL"

gene

8729..9319

/gene="mog"

/locus\_tag="t0008"

8729..9319

/locus\_tag="t0008"

/note="corresponds to STY0008 from Accession AL513382"

Salmonella typhi CT18"

/codon\_start=1

/transl\_table=11

/product="molybdopter in biosynthesis Mog protein"

/protein\_id="AA067742.1"

/db\_xref="GI:29136175"

/translation="MNTLRIGLVSIDRASSGVVODKIGIPALBEWLASALTTPPELOT  
RLIPDEQALTEQCLVDEMSELVLTGTGTGPARDVTPDRLATADREMFGFEGQ  
MRQISLRFVFTALISRQGVIRKQALINLPQPKSIKTELEGVKAODGDSVSPFGIPA  
SVYPCQLDGPVETAPVEVAAFRPKSARRENSS"

gene

/gene="yaaH"

/locus\_tag="t0009"

complement(9376..9942)

/gene="yaaH"

complement(9376..9942)

/locus\_tag="t0009"

/note="corresponds to STY0009 from Accession AL513382"

Salmonella typhi CT18"

/codon\_start=1

## Alignment Scores:

Pred. No.: 2,27e-53 Length: 300169  
Score: 990.00 Matches: 233  
Percent Similarity: 64.18% Conservative: 59  
Best Local Similarity: 51.21% Mismatches: 139  
Query Match: 43.81% Indels: 24  
DB: 1 Gaps: 10

US-09-701-229-2 (1-448) x AE016834 (1-300169)

Qy 2 SerLeuileAlaSerAspHisPheArg-----IleValValGlyLeuGlyLysSer 18  
Db 148797 ACGTAATCATGGCTGATTAACAGGACAAAACCTGCTATTATCGGTCTGGGCTTAAC 148856  
Qy 19 GlyMetSerLeuValArgTyrLeuAlaArgGlyLeuProPheAlaValAlaAspThr 38  
Db 148857 GGAATCTCTGCTGGACTTTTCTCGCCCGCGCGTACCGCGCGGTGATGATACT 148916  
Qy .39 ArgGluAsnProProGluLeuAlaThrLeuAlaGlnTyrProGln----ValGluVal 57  
||| ||||| ||| ||| ||||| |||||

Db 148917 CGCGTGCAGCGCGCGGTCTGGATAAGCTG-----CCGCAAGAGGTTGAGCGT 148964  
Qy 58 ArgCysGlyGluLeuAspAlaGluPheLeuCysSerAlaArgGluLeuTyrValSerPro 77  
Db 148965 CACGTTGGCGCCCTGAACGACGAGTGGCTCTTACGGCGGATTTAATGCTGCCAGCCCT 149024  
Qy 78 GlyLeuSerLeuArgThrProAlaLeuValGlnAlaAlaLysGlyValArgLysSer 97  
Db 149025 GGTATTGCTCTGGCGCATCCCTCTTACGCTGCGCGACGCTGGAGTGAATATGTC 149084  
Qy 98 GlyAspIleAspLeuPheAlaArgGluAlaLysAlaProIleValAlaIleThrGlySer 117  
Db 149085 GCGCATATCGAACTGTTTCCCGGAAACGCGGCGGATTTGGCCATCACCAGGCTCG 149144  
Qy 118 AsnAlaLysSerThrValThrLeuValGlyGluMetAlaValAlaAlaAspLysArg 137  
Db 149145 AACGGCAAAAGCAGCCGTGACCCCTTAGTGGCGAGATGGCGAAGCGCGGCGTCAAT 149204  
Qy 138 ValAlaValGlyGlyAsnLeuGlyThrProAlaLeuAspLeuLeuAlaAspIleGlu 157  
Db 149205 GTCGGCGTGGCGCAATATCGGTCTTCCGCGCTGATGCTGCTGGATGCCGATCGCGAA 149264  
Qy 158 LeuTyrValLeuGluLeuSerSerPheGlnLeuGluThrCysAspArgLeuAsnAlaGlu 177  
Db 149265 TTGTACGTGCTGGAATATCCAGTTTCAATTTGGAACCTACCTCAAGTTTGCACAGCGCG 149324  
Qy 178 ValAlaThrValLeuAsnValSerGluAspHisMetAspArgTyrAsp---GlyMetAla 196  
Db 149325 CGGCAACGGTGTCAACGTCACGTGAAGATCATATGACCGGTATCCGTTTGGTTGGAA 149384  
Qy 197 AspTyrHisLeuAlaLysHisArgIlePheArgGlyAlaArgGlyGlnValValAsnArg 216  
Db 149385 CAGTACCGCGCGCGGAAACTGCGCGTCTACGAGAAGCGGAAGTGTGCTAGTGAATGCC 149444  
Qy 217 AlaAspAlaLeuThrArgProLeuIleAlaAspThrValProCysTyrSerPheGlyLeu 236  
Db 149445 GATGACGCGTTGACGATGCGGTACGCGGCGCGTACGCGCGTCTGCTGAGTTTGGCGTC 149504  
Qy 237 AsnLysProAspPheLysAlaPheGlyLeuIleGluGluAspGlyGlnLysTrpLeuAla 256  
Db 149505 AATATGGTGATTAATCAC-----CTTAATCGTCAGAGGCGGGAACCTGCTACGA 149555  
Qy 257 PheGlnPheAspLysLeuLeuProValGlyLeuLysIleArgGlyAlaHisAsnTyr 276  
Db 149556 GTCAAGAGGTGAGAAGTGTCAACGTGAAGAGAGATGAAGCTTCCGCTCAGCATAACTAT 149615  
Qy 277 SerAsnAlaLeuAlaAlaLeuAlaLeuGlyHisAlaValGlyLeuProPheAspAlaMet 296  
Db 149616 ACCAATCGTTAGCGCGCTGGCGTCCGCGATCCGCTAGTCTGCCCGCGCCAGCAGT 149675  
Qy 297 LeuGlyAlaLeuLysAlaPheSerGlyLeuAlaHisArgCysGlnTyrPvalArgGluArg 316  
Db 149676 TTGAAGCGTTGACGACATTTTACCGGTCTGGCGCATCGTTCACGTGGCGTTGGAGCATT 149735  
Qy 317 GlnGlyValSerTyrAspAspSerLysAlaThrAsnValGlyAlaAlaLeuAlaAla 336  
Db 149736 AACGCGCTCGTTGGATCAACGACTCGAAACGACCAATGTCCGAGTACCGAAGCGCG 149795  
Qy 337 IleGluGlyLeuGlyAlaAspIleAspGlyLysLeuValLeuLeuAlaGlyLysGly 356  
Db 149796 CTAACCGTTTTG-----CATGTGGAGCGGTACGCTGCTGCTGCTGGCGCGCGCGGT 149849  
Qy 357 LysGlyAlaAspPheHisAspLeuArgGluProValAlaArgPheCysArgAla----- 374  
Db 149850 AAGTCGCGAGACTTT-----TCTCCGCTGGCGCGCTATCTGACCGCGGATCGT 149897  
Qy 375 -----ValValLeuLeuGlyArgAspAlaGlyLeuIleAlaGlnAlaLeuGlyAsnAla 392  
Db 149898 ATCCGCGCTGATTGCTTTGGCGCGATGGCGCGCAGCTTGGC---GCACCTCGCTCGGAA 149954  
Qy 393 ValProLeuValArgValAlaThrLeuAspGluAlaValArgGlnAlaAlaGluLeuAla 412  
Db 149955 ATC---GCCCAACAGACTGAGACGATGGAAGCGGATGCGCTTTGCTGCTGCCCGCCGCGT 150011



```

/transl_table=11
/product="2-isopropylmalate synthase"
/protein_id="AAL19077.1"
/db_xref="GI:16418612"
/translation="MSQOVIIFDTTLRDGEQALQASLSAKEKIQIALALERMGVYME
VGPVSPGDFESVQIARTIKNSRVCALARCVEKDIDVAQAQKLVADAFRIHTFIAT
SPMHIATKRLSTLDEIVIERAVYMKRARNYTDVEFSCEDAGRTPPVDGLARVVEAAIN
AGARTINIDPTGYTTPFFAGIISGLYERVPNIDKAIISVHTHDDGLTAGVNSLAAY
HAGAROVGAMNGIGRAGNCALEEVIMAKVRKIDMNVHTNINHHIEIWRTSQTSOI
CNMPIPANAIYVSGSAFASGGIHODGVLKRENEYIMTPESIGLNOIQNLTSRSGR
AAYKRMEMGKDTYINMDHLYDAELKLADKKQGVFDLLEALAIINKQEEPEHFR
LDYFVQSSSDIATASVLAAGEEIKAEAAANGPVDIAIYQAINRITGYDELVKYD
LNAGKGKDALGQDVIYVNHGRFRFHVGVLATDIVESSAKAMVHVLNNIWRAEVEKE
LQRAQKNENKETY"
RBS
complement(3417..3422)
/genes="leuB"
/notes="putative RBS for leuB; RegulonDB: STMS1H000492"
complement(4994..4999)
/genes="leuA"
/notes="putative RBS for leuA; RegulonDB: STMS1H000493"
complement(5067..5298)
/genes="leuL"
/notes="synonym: STM0114"
complement(5067..5153)
/genes="leuL"
/notes="leu operon leader peptide. (SW:LPL_SALTY)"
/codon_start=1
/transl_table=11
/product="leu operon leader peptide"
/protein_id="AAL19078.1"
/db_xref="GI:16418613"
/translation="MSHIVRFTGLLLNFAFIVGRVPVGGIQH"
complement(5162..5167)
/genes="leuL"
/notes="putative RBS for leuL; RegulonDB: STMS1H000494"
complement(5271..5279)
/genes="leuL"
/notes="putative -10_signal for leuL;
RegulonDB: STM1TH004472"
complement(5290..5298)
/genes="leuL"
/notes="putative -35_signal for leuL;
RegulonDB: STM1TH004472"
5798..5754
/genes="leuO"
/notes="synonym: STM0115"
5798..5803
/genes="leuO"
/notes="putative RBS for leuO; RegulonDB: STMS1H000495"
5810..5754
/genes="leuO"
/notes="probable activator protein in leuabcd operon.
(SW:LEUO_SALTY)"
/codon_start=1
/transl_table=11
/product="putative Lysr family transcriptional regulator"
/protein_id="AAL19079.1"
/db_xref="GI:16418614"
/translation="MSEVTEKPHLLDMGKPOLRMVDLNLITVFDVMOEONITRAAH
TLGMSPAVSNVARKLVMPNDELFRYRGIOPTARAFQIFGFSVQAOLQVONELPG
SGEPTSSSRVNLVCSPLDNLITSOYNRVEKIAPNIHVFKLSNLQNTQEHOLRYQ
ETEFVTSYEEFRPEFTSPVLFKDEMVLASRKRPISGFLPEGDVYNQEHAVYSLDR
YASFQSPWDTDPKQSVAYQGMALISVLNVQSOTHLVAIPRWLAEEFAESLDLIQL
PLPLKLSRTECYLSWHEAAGRDKGQWMDLLVSVCKR"
6941..6953
protein_bind
1.55e-54 Length: 22348
Pred. No.: 987.00 Matches: 232
Score: 63.96% Conservative: 59
Percent Similarity: 50.99% Mismatches: 140
Best Local Similarity: 43.67% Indels: 24
Query Match: 1
DB: 1

```

```

US-09-701-229-2 (1-448) x AE008699 (1-22348)
QY 2 SerLeuIlealaSerAspHisPheArg-----IleValValGlyLeuGlyLysSer 18
Db 18632 ACCTTAATCATGCTGATTACCAAGCAAAAGCTGCTATTATGCTTGGCTTAAC 18691
QY 19 GlyMetSerLeuValArgTyrLeuAlaArgGlyLeuProPheAlaValAlaThr 38
Db 18692 GGACTCTCTGCGTGGNCTTTTCTCGCCCGGGGTGACGCCGGGTGATGGATACT 18751
QY 39 ArgGluAsnProProGluLeuAlaThrLeuArgAlaGlnTyrProGln----ValGluVal 57
Db 18752 CGCGTCACGCCCGCGGTCTGGATAAGCTG-----CGCAAGAGGTGTAGCGT 18799
QY 58 ArgCysGlyLeuLeuAspAlaGluPheLeuCysSerAlaArgGluLeuTyrValSerPro 77
Db 18800 CAGTTTGGCGGCTGAACGACGAGTGGCTTTAGCGCGGATTAATGTCGCCGAGCCT 18859
QY 78 GlyLeuSerLeuArgThrProAlaLeuValGlnAlaAlaAlaLysGlyValArgIleSer 97
Db 18860 GGTATTGCTCTGGCGCATCCCTCTCTTAGCGCTGCCCGCAGCGCTGGAGTGTGTC 18919
QY 98 GlyAspIleAspLeuPheAlaArgGluAlaLysAlaProIleValAlaIleThrGlySer 117
Db 18920 GCGGATATCGAAGTGTTCGCGAAGCGCGGATTTGCGCATCACCGGCTCG 18979
QY 118 AsnAlaLysSerThrValThrThrLeuValGlyGluMetAlaValAlaAlaAspLysArg 137
Db 18980 AACGGCAAAAGCACCGTGCACCTTAGTGGCGGAGATGCGAAAGCGCGGCGTCAAT 19039
QY 138 ValAlaValGlyGlyAsnLeuGlyThrProAlaLeuAspLeuLeuAlaAspIleGlu 157
Db 19040 GTCGGCGTTGGCGCAATATCGCTTCGCGCGCTGATGCTGGTGGATCGCGGAA 19099
QY 158 LeuTyrValLeuGluLeuSerSerPheGlnLeuGluThrCysAspArgLeuAsnAlaGlu 177
Db 19100 TTGTAGCTGCTGGAAATATCCAGTTTCCAACTGGAACACTACCTCAAGTTTTCGAA 19159
QY 178 ValAlaThrValLeuAsnValSerGluAspHisMetAspArgTyrAsp---GlyMetAla 196
Db 19160 GCGGCAACGGTGTCAACGTCATGAAGATCATATGACCGGATTCGCTTTGGTTGCCAA 19219
QY 197 AspTyrHisLeuAlaLysHisArgIlePheArgGlyAlaArgGlnValValAlaAsnArg 216
Db 19220 CAGTACGCGGCGCGAAACTGCGCTCTACGAGAGCGCAAGTGTCCGTAGTGAATGCC 19279
QY 217 AlaAspAlaLeuThrArgProLeuIleAlaAspThrValProCysTyrSerPheGlyLeu 236
Db 19280 GATGACGCGTTGACGATGCGGTCACGCGGCGCGATGAACGCTCGCTAGCTTTGGCGTC 19339
QY 237 AsnLysProAspPheLysAlaPheGlyLeuIleGluAspGlyGlnLysTyrLeuAla 256
Db 19340 AATATGGCGCATATCAC-----CTTAATCGTCAGCGCGGCAAACTGGCTCGGA 19390
QY 257 PheGlnPheAspLysLeuLeuProValGlyGluLeuLysIleArgGlyAlaHisAsnTyr 276
Db 19391 GTCAAGCGCGAGAGGTGCTGAACGTGAAGAGAGATGAAGCTTTCCGGTCAACATACTAT 19450
QY 277 SerAsnAlaLeuAlaAlaLeuAlaLeuGlyHisAlaValGlyLeuProPheAspAlaMet 296
Db 19451 ACCAATGCGTTAGCGGCGTGGCGGCTGCGGATGCCGTAGCTGCTCGCGCGCAGCAGT 19510
QY 297 LeuGlyAlaLeuLysAlaPheSerGlyLeuAlaHisArgCysGlnTyrPvalArgGluArg 316
Db 19511 TTGAAGCGGTGACGACATTTACCGGTCTGGCATTCGCTTCAGCTGGCGGTGGGACAT 19570
QY 317 GlnGlyValSerTyrTyrAspAspSerLysAlaThrAsnValGlyAlaAlaLeuAlaAla 336
Db 19571 AACGGCGTACGTTGGATCAACGACTCGAAAGCGCAACACGTCGCGCAGTACCGAGCGCG 19630
QY 337 IleGluGlyLeuGlyAlaAspIleAspGlyLysLeuValLeuLeuAlaGlyAspGly 356

```

```

Db      19631 CTAACGGTTG-----CATGTGGAGGTACCTACATCTGCTGCTGGCGGCGGACGGT 19684
QY      357 LysGlyAlaAspPheHisAspLeuArgGluProValAlaArgPheCysArgAla----- 374
Db      19685 AAGTCGGCAGACTTT-----TCTCCGCTGACACGGCTATCTGACCGCGGATCGT 19732
QY      375 -----ValValLeuLeuGlyArgAspAlaGlyLeuIleAlaGlnAlaLeuGlyAsnAla 392
Db      19733 ATCCGCTGTATTGCTTTGGCGCGATGGCGCAGCTTGC-----GCACTGGCTCCGGA 19789
QY      393 ValProLeuValArgValAlaThrLeuAspGluAlaValArgGlnAlaGluLeuAla 412
Db      19790 ATC---GCCACAGACTGAGCAGTAGGAGAGCGATCGTTGCTGGCGCGCGCGTT 19846
QY      413 ArgGluGlyAspAlaValLeuLeuSerProAlaCysAlaSerLeuAspMetPheLysAsn 432
Db      19847 CAGCCGGGTGATAGTGTGCTGTCGCGCGCTGGCTAGCTCGATCAGTTTAAANAT 19906
QY      433 PheGluGlyArgLeuPheAlaLysAlaValGluLeu 447
Db      19907 TTGAGCAACGGGGGATGCTTTACCCGCTGCTGGCGAAGGAGTTA 19951

-RESULT 9
ECMUROY      2608 bp      DNA      linear      BCT 12-SEP-1993
LOCUS      Escherichia coli murD gene and ORF-Y (EC 6.3.2.9).
DEFINITION      X51584.X52117
ACCESSION      X51584.1 GI:42058
VERSION      UDP-N-acetylmuramoyl-L-alanyl-D-glutamate synthetase; unidentified
KEYWORDS      reading frame.

SOURCE      Escherichia coli
ORGANISM      Escherichia coli
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; Escherichia.
REFERENCE      1 (bases 1 to 2608)
AUTHORS      Ikeda,M., Wachi,M., Ishino,F. and Matsuhashi,M.
TITLE      Nucleotide sequence involving murD and an open reading frame ORF-Y
JOURNAL      spacing murF and ftsW in Escherichia coli
MEDLINE      Nucleic Acids Res. 18 (4), 1058 (1990)
PUBMED      90192099
REFERENCE      2 (bases 1 to 2608)
AUTHORS      Ikeda,M.
TITLE      Direct Submission
JOURNAL      Submitted (02-JAN-1989) Ikeda M., Institute of Applied
Microbiology, The University of Tokyo, 1-1-1 Yayoi, Bunkyo-ku,
Tokyo 113, Japan
REFERENCE      3 (bases 26 to 1354)
AUTHORS      Flouret,B.
TITLE      Direct Submission
JOURNAL      Submitted (15-MAR-1990) Flouret B., URA 1131 du Centre National de
la Recherche Scientifique, C N R S, Biochimie Molculaire et
Cellulaire, Batiment 432 Universite Paris-Sud, 91405 Orsay, France
COMMENT      Sequence independently determined by [3], but authors accept
sequence [1] as definitive.
FEATURES      Location/Qualifiers
source      1..2608
            /organism="Escherichia coli"
            /mol_type="genomic DNA"
            /db_xref="taxon:562"
            /map="2min"
            /gap="2min"
            /note="pot. ribosome binding site"
            111..1193
            /note="ORF-Y (AA 1-360)"
            /codon_start=1
            /transl_table=11
            /protein_id="CAA35932.1"
            /db_xref="GI:42059"
            /db_xref="SWISS-PROT:P15876"
            /translation="MLVLAHLVKYISYGFNNVFSYLTFRVLSLLTALFLISLMGPRM
            1AHLQKLSFGQVRNDGPESHFRKGTPTMGIMLITAIIVISLLWAYPSNPYVCVL

```

```

VVLVGVGVIGVDDYKVRKDKTGLIARKYFWMVSVIALGVAFALYLAGKDTPTATQL
VVPFFKDVMPQLGLFYILLAYFVIVGTGNVNLDTGLDLAIPTVFAAGGALVAMA
TGNMNFASVYHPIYLRHAGELIVCTAIVGAGLGLFWNTYPAQVFMGVGSLALGGA
LGIIVLLRQEFLLVIMGVFVETLSVLQVSEFKLRQIRFMAPIHHHVELKGWP
EPRIVRFWIISILMLVLIGLATLKVR"
1018..1023
/note="pot. -35 region"
1043..1048
/note="pot. -10 region"
1181..1185
/note="pot. ribosome binding site"
1196..12512
/note="murD gene product (AA 1-438)"
/codon_start=1
/transl_table=11
/protein_id="CAA35933.1"
/db_xref="GI:42060"
/db_xref="SWISS-PROT:P14900"
/translation="MADYQGNVVIIGLGLTGLSCVDFPLARGVTPRVMDTRMTPPGL
DKLPEAVERHTGSLNDEWLMADLIVASPGIALHSPLSAADAAGIEIVGDIELFCRE
AQAPIVAITGSKSTVTLVGEAKAAAGVNGVNGIGLPAIMLLDDCELYLELS
SFOLETTSSLAQAATILNVTEDHMDRYPFGLQOYRAAKRLIYENAKVCVWADALY
MPIRGADERCVSFGVNMGDYHLNHQOGETWLRVKEKLVNVEKLSGHNVTNALAA
LALADAAGLPRASSLSKALTFTTGLPHREYVLEHNGVRHINDSKATNVGSTAALNGL
HYDGTLLHLLGGDKSADFSPILRYLNGDNRVLYCFGRDGAQLAALRVEVAQTETME
QAMRLIAPRVQPGDMVLVSPACASLDQFNKFEQGNFAFLAKELG"
BASE COUNT      530 a      631 c      777 g      670 t
ORIGIN

Alignment Scores:
Pred. No.:      1,35e-55      Length:      2608
Score:      986.00      Matches:      230
Percent Similarity:      63.96%      Conservative:      61
Best Local Similarity:      50.55%      Mismatches:      140
Query Match:      43.63%      Indels:      24
DB:      1      Gaps:      10

US-09-701-229-2 (1-448) x ECMUROY (1-2608)
QY      2 SerLeuIleAlaSerAspHisPheArg-----IleValValGlyLeuGlyLysSer 18
Db      1187 AGCTTAATCATGCTGCTGATTATCAGGGTAAATAATGCTGCTATATCGGCTCGGCTCACC 1246
QY      19 GlyMetSerLeuValArgTyrLeuAlaArgGlyLeuProPheAlaValValAspThr 38
Db      1247 GSGCTTTCGCTGCTGAGCTTTTTCCTCGCTCGCTGTCGCGCGCTGATGATGATG 1306
QY      39 ArgGluAsnProProGluLeuAlaThrLeuArgAlaGlnTyrProGln---ValGluVal 57
Db      1307 CGTATGACACCCCTCGGCTGGATAAATTA-----CCGGAAGCCCTGAAGACG 1354
QY      58 ArgCysGlyGluLeuAlaGluPheLeuLysSerAlaArgGluLeuTyrValSerPro 77
Db      1355 CACACGGCAGCTCTCAATGATGAATGGCTGATGGCGGAGATCTGATGTCGCCAGTCCC 1414
QY      78 GlyLeuSerLeuArgThrProAlaLeuValGlnAlaAlaLysGlyValArgIleSer 97
Db      1415 GGTATTGCACCTGGCGCATCCATCCCTTAAGCGCTGCGCTGATGCCGAATCGAATCGTT 1474
QY      98 GlyAspIleAspLeuPheAlaArgGluAlaLysAlaProIleValAlaIleThrGlySer 117
Db      1475 GCGGATATCGAGCTGTTCTGCGCGAAGCACACCGCATGTTGCGCATATACCGGTTCT 1534
QY      118 AsnAlaLysSerThrValThrThrLeuValGlyGluMetAlaValAlaAlaAspLysArg 137
Db      1535 AACGCGCAANAGCAGCTCACCACGCTAGTGGTGAATGGCGAAGCGCGGGGTTAAC 1594
QY      138 ValAlaValGlyGlyAsnLeuGlyThrProAlaLeuAspLeuLeuAlaAspIleGlu 157
Db      1595 GTTGTGTGGGTGGCAATATTGGCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1654
QY      158 LeuTyrValLeuGluLeuSerSerPheGlnLeuGluThrCysAspArgLeuAsnAlaGlu 177

```





```

/note="factor Sigma70; promoter leu; documented +1
at83735"
protein_bind
complement(201..221)
/note="No predicted promoter"
/bound_moiety="Lexa predicted site"
protein_bind
complement(201..221)
/note="central position to predicted promoter: -0.5"
/bound_moiety="Lexa predicted site"
protein_bind
complement(354..368)
/note="No predicted promoter"
/bound_moiety="Lrp documented site"
protein_bind
complement(354..368)
/note="central position to leuABCD promoter: -156.5"
/bound_moiety="Lrp documented site"
protein_bind
460..474
/note="central position to predicted promoter: -28.5"
/bound_moiety="Lrp predicted site"
promoter
462..490
/note="factor Sigma70; predicted +1 start at 84042"
complement(519..537)
protein_bind
/note="central position to predicted promoter: -319.5"
/bound_moiety="TyrR predicted site"
protein_bind
554..565
/note="central position to predicted promoter:64"
/bound_moiety="AraC predicted site"
gene
646..1767
/note="leuO"
/note="synonym: b0076"
CDS
646..1767
/note="leuO"
/function="putative regulator; Amino acid biosynthesis:
Leucine"
/note="o373; 100 pct identical to LEUO_ECOLI SW:
p10151(290 aa) but contains 59 additional N-ter aa and 24
C-term residues; 100 pct identical to PIR: S40589 but
contains 24 additional C-term residues"
/codon_start=1
/transl_table=11
/product="probable transcriptional activator for leuABCD
operon"
/protein_id="AAC73187.1"
/db_xref="GI:1786264"
/translation="MTHSTAMDSVFIRIFMFSEFYSCFFLFYMHDKSYSSGLFIC
IPRERELSVUELSPVEQTDHPHTELSKPLQMLVDNLITVDVAMQONITRAA
HYLGSPAPNAVARKVMENDELFRYGQIPTARAFQFGSRQALQVQNELP
GSGFSPASERVHLCVCSPLDSILTSIQIYHIEQIAENIHVMFKSSLNQTEHLRY
QETFEVSYEDFHRPFTSVPLFKDEMVLVASKNHPTIKGPKLKHVDVNEQHAASLD
RFASQSPWYTDVKOASIAVQGMAMSVLSVSVQTHLVATAPRWLAEEFAESLELQV
LPLPLKNSRRCYLSWHEAGRDKGHWERQLYSICKR"
protein_bind
1802..1816
/note="central position to predicted promoter: -39.5"
/bound_moiety="Lrp predicted site"
promoter
1813..1839
/note="factor Sigma70; promoter livIhp4; documented+1 at
85394"
protein_bind
1833..1847
/note="central position to predicted promoter: -8.5"
/bound_moiety="Lrp predicted site"
promoter
1838..1866
/note="factor Sigma70; promoter livIhp3; documented+1 at
85420"
protein_bind
1915..1929
/note="central position to predicted promoter:73.5"
/bound_moiety="Lrp predicted site"
protein_bind
1949..1963
/note="central position to predicted promoter:107.5"
/bound_moiety="Lrp predicted site"
promoter
1955..1984
/note="factor Sigma70; promoter livIhp2; documented+1 at
85534"
protein_bind
1978..1992
/note="central position to predicted promoter:136.5"
/bound_moiety="Lrp predicted site"

1995..3809
/gene="ilvI"
/note="synonym: b0077"
CDS
1995..3809
/gene="ilvI"
/EC_number="4.1.3.18"
/function="enzyme; Amino acid biosynthesis: Isoleucine,
Valine"
/note="o604; 98 pct identical (1 gap) to 522 residues
fromILVI_ECOLI SW: P00893 (566 aa)"
/codon_start=1
/transl_table=11
/product="acetolactate synthase III, valine sensitive,
large subunit"
/protein_id="AAC73188.1"
/db_xref="GI:1786265"
/translation="MFMSLAANCLSKIGRLMCFTHFFRQTVRQANEMLSGAEVVRSL
IDQGVKVFYGGPGGAVLDYDALHTVGGIDHVLVRHEQAAMHADGLARAGEVGVVL
VTSFGPATNATIGATAYMDSIPLVLVSGVATSLIGDYAFQECMDVGISRPVVKHSF
LVKQTEDIPQVLKKAFLAASGPGPVVDLPKDLINPANKLPYVWPESVMSRNPNT
TTGHGQIKRALQTLVAAKPVVYVGGGATAGCHQOLKETVEALNLPVCSLMGLGA
FPATHROALGMLGMHGTVEANNTMHNADVIFAVGVRFDDRTTNLAKYCPNATLHID
IDPISISKTVTADIPVGDARQVLEQMLELSSQSAHQPLDEIRDWMQOIQWRARQC
LKYDTHSEKIRKQAVIETLWRLTKGDAYTSDVGQHQMFALYIPFDPKRWNNSGGL
GTMGFGLPALGVKMALEPTEVVCVTGDSIQMNIQELSTALQVELPVLVNNRRL
GMVKQWQDMYISGRHSYSQMSLPDFVRLAAYGHVGIQISHPELESKLSLEAEQVR
NNRLVFVDVTVDGSEHYVPMQIRGGGMDMWSKTERT"
1998..2012
/gene="ilvI"
/note="central position to predicted promoter:156.5"
/bound_moiety="Lrp predicted site"
2017..2045
/promoter
2017..2045
/note="factor Sigma70; promoter livIhp1; documented+1 at
85597"
gene
3812..4303
/note="ilvH"
/note="synonym: b0078"
CDS
3812..4303
/note="ilvH"
/EC_number="4.1.3.18"

Alignment Scores:
Pred. No.: 1.74e-54 Length: 21757
Score: 986.00 Matches: 230
Percent Similarity: 63.96% Conservative: 61
Best Local Similarity: 50.55% Mismatches: 140
Query Match: 43.63% Indels: 24
DB: 1 Gaps: 10

US-09-701-229-2 (1-448) x AE000118 (1-21757)
Qy 2 SerLeuileAaSerAspHisPheArg-----ileValValGlyLeuGlyLysSer 18
Db 13533 ACGTTAATCATGGCTGATTATCAGGGTAAATAATGTCGTCATTATCGGCTGGGCTCAC 13592
Qy 19 GlyMetSerLeuValArgTyrLeuAalaArgGlyLeuProPheAaValValAspPhe 38
Db 13593 GGGCTTTCCTCGTGGGACTTTTCTCGCTCGCGGTGACGCGCGCTTATGGATACG 13652
Qy 39 ArgGluAsnProProGluLeuAalaThrLeuAalaGlnTyrProGln---ValGluVal 57
Db 13653 CGTATGACACCGCTGGCTGGATAAATTA-----CCGAAGCGGTAGAACGC 13700
Qy 58 ArgCysGlyGluLeuAspAalaGluPheLeuCysSerAaArgGluLeuTyrValSerPro 77
Db 13701 CACACGGGCAGCTCTGAATGATGATGCTGATGGCGGCGAGATGATGTCGCCAGTCCC 13760
Qy 78 GlyLeuSerLeuArgThrProAalaLeuValGlnAalaAalaLysGlyValArgIleSer 97
Db 13761 GGTATTGCACGTGGCGCATCCATCCTTAAGCGCTGCGCGTATGCGGAATCGAAATCGTT 13820
Qy 98 GlyAspIleAspLeuPheAalaArgGluAalaLysAlaProIleValAalaIleThrGlySer 117

```



```

Db      14196  CAGTATCGTCGACGAAACCTGCGCATTTACGAAACGCGAAAGTTTGCCTGCTTAATGCT 14255
QY      217   AlaAspAlaLeuThrArgProLeuIleAlaAspThrValProCysTrpSerPheGlyLeu 236
Db      14256  GATGATGCCCTTACCAATGCCGATTCGGTGGCGGATGAACGCTGCGTCAGCTTGGCGTC 14315
QY      237   AsnLysProAspPheLysAlaPheGlyLeuIleGluGluAspGlycInLysTrpLeuAla 256
Db      14316  AACATGGGTGACTATCAC-----CTGAATCATCAGCAGCGCGAAACCTGGCTGCGG 14366
QY      257   PheGlnPheAspLysLeuLeuProValGlyGluLeuLysIleArgGlyAlaHisAsnTrp 276
Db      14367  GTTAAAGCGCGAGAAAGTGTGAATGTAAGAGATGAACACTTTCGGCGCAGCATAACTAC 14426
QY      277   SerAsnAlaLeuAlaLeuLeuGlyHisAlaValGlyLeuProPheAspAlaMet 296
Db      14427  ACCAATGCCCTGGCGCGTGGCGTGGCGAGATGTCAGAGGTTCACGGTCCCGCAGCAGC 14486
QY      297   LeuGlyAlaLeuLysAlaPheSerGlyLeuAlaHisArgCysGlnTrpValArgGluArg 316
Db      14487  CTGAAGCGCTTAACACACTTCACTGCTGTCGCCATCGCTTTGAAGTTGCTGCTGGAGCAT 14546
QY      317   GlnGlyValSerTyrThrAspSerLysAlaThrAsnValGlyLeuAlaLeuAla 336
Db      14547  AACGGCGTACGTTGGATTACGATTGCAAGACGACCAACGTCGCGCAGTACGGAACGGCG 14606
QY      337   IleGluGlyLeuGlyAlaAspIleAspGlyLysLeuValLeuAlaGlyLysAspGly 356
Db      14607  CTGAATGGCGCTG-----CACGTAGACGCGACACTGCATTGTTGCTGGTGGCGCATGGT 14660
QY      357   LysGlyAlaAspPheHisAspLeuArgGluProValAlaAlaArgPheCysArgAla----- 374
Db      14661  AAATCGCGCGACTTT-----AGCCCACTGGCGCGTTACCTGAATGGCGATAAC 14708
QY      375   -----ValValLeuLeuGlyArgAspAlaGlyLeuIleAlaGlnAlaLeuGlyAsnAla 392
Db      14709  GTACGTCGTATGTTGCTGTCGCGCGCGCAGCTGGCG-----GCGCTACGCCCGGAA 14765
QY      393   ValProLeuValArgValAlaThrLeuAspGluAlaValAlaValArgGlnAlaGluLeuAla 412
Db      14766  GTG---GCAGAACAAACCGAACTATGGAACAGCGATGCGCTGCTGCTCCGCGTGT 14822
QY      413   ArgGluGlyAspAlaValLeuLeuSerProAlaCysAlaSerLeuAspMetPheLysAsn 432
Db      14823  CAGCGCGCGGATATGTTGCTCTCCCGAGCTGTGCCAGCTTGCATCAGTTCAAGAAC 14882
QY      433   PheGluGluArgGlyArgLeuPheAlaLysAlaValGluGluLeu 447
Db      14883  TTTGAACACGAGGCAATGATGTTGCCCGCTGCGGAGGAGTTA 14927

RESULT 12
EC2MIN
ID      EC2MIN      standard; DNA; PRO; 28277 BP.
XX      X55034; M10429;
AC      X55034;
SV      X55034.1
XX      21-FEB-1991 (Rel. 27, Created)
DT      05-JUL-1999 (Rel. 60, Last updated, Version 31)
XX      E. coli 2 minute region
DE      ddl gene; envA gene; ftsA gene; ftsQ gene; ftsW gene; ftsZ gene; ilvH gene;
XX      ilvI gene; leuA gene; leuO gene; mray gene; murC gene; murD gene;
KW      mreE gene; murF gene; murG gene; mutR gene; orfA; orfB; orfC; orfX;
KW      pbpB gene; secA gene; shl gene.
OS      Escherichia coli
OC      Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC      Enterobacteriaceae; Escherichia.
XX

```

[1] 1-28277  
Ayala J.A.;  
Submitted (08-JAN-1991) to the EMBL/GenBank/DBJ databases.  
Ayala J.A., Instituto de Biología Molecular, Centro de Biología Molecular,  
Universidad Autónoma, Canto-Blanco 28049, Madrid, Spain.

[2] Ayala J.A.;  
"Regulation of transcription at the 2-minute region of the genetic map of  
Escherichia coli";  
Unpublished.

[3] 1-306  
MEDLINE; 82078077.  
PUBMED; 6171647.  
Wessler S.R., Calvo J.M.;  
"Control of leu operon expression in Escherichia coli by a transcription  
attenuation mechanism";  
J. Mol. Biol. 149(4):579-597(1981).

[4] 268-1130  
MEDLINE; 86223773.  
PUBMED; 3519576.  
Haughn G.W., Wessler S.R., Gemmill R.M., Calvo J.M.;  
"High A + T content conserved in DNA sequences upstream of leuABCD in  
Escherichia coli and Salmonella typhimurium";  
J. Bacteriol. 166(3):1113-1117(1986).

[5] 843-1812  
MEDLINE; 88320486.  
PUBMED; 3413113.  
Henikoff S., Haughn G.W., Calvo J.M., Wallace J.C.;  
"A large family of bacterial activator proteins";  
Proc. Natl. Acad. Sci. U.S.A. 85(18):6602-6606(1988).

[6] 1799-2187  
MEDLINE; 85234358.  
PUBMED; 3891724.  
Haughn G.W., Squires C.H., Defelice M., Largo C.T., Calvo J.M.;  
"Unusual organization of the ilvIH promoter of Escherichia coli";  
J. Bacteriol. 163(1):186-198(1985).

[7] 2101-4431  
MEDLINE; 83272971.  
PUBMED; 6308579.  
Squires C.H., Defelice M., Devereux J., Calvo J.M.;  
"Molecular structure of ilvIH and its evolutionary relationship to ilvG in  
Escherichia coli";  
Nucleic Acids Res. 11(15):5299-5313(1983).

[8] 4274-6093  
MEDLINE; 90330585.  
PUBMED; 2198273.  
Leclerc G., Noel G., Drapeau G.;  
"Molecular cloning, nucleotide sequence and expression of shl, a new gene  
in the 2-minute region of the genetic map of Escherichia coli";  
J. Bacteriol. 172(8):4696-4700(1990).

[9] 6088-7587  
MEDLINE; 90251464.  
Gomez M.J., Fluoret B., Van Heijenoort J., Ayala J.A.;  
"Nucleotide sequence of the regulatory region of pbpB gene of Escherichia  
coli";  
Nucleic Acids Res. 18:2813-2813(1990).

XX [10] MEDLINE; 90124047.  
RN 7316-10074  
RP MEDLINE; 83296957.  
RX PUBMED; 6350821.  
RA Nakamura M., Maruyama I.N., Soma M., Kato J.I., Suzuki H., Hirota Y.;  
RT "On the process of cellular division in *Escherichia coli*: nucleotide  
RL sequence of the gene for penicillin-binding protein 3.";  
RL Mol. Gen. Genet. 191(1):1-9(1983).  
XX [11]  
RN MEDLINE; 90124047.  
RX PUBMED; 2692800.  
RA Tao JS., Ishiguro E.E.;  
RT "Nucleotide sequence of the murE gene of *Escherichia coli*";  
RL Can. J. Microbiol. 35(11):1051-1054(1989).  
XX [12]  
RN MEDLINE; 90328986.  
RX PUBMED; 2198024.  
RA Michaud C., Parquet C., Flouret B., Blanot D., Van Heijenoort J.;  
RT "Revised interpretation of the sequence containing the murE gene encoding  
RL the UDP-N-acetylmuramyl-tripeptide synthetase of *Escherichia coli*";  
RL Biochem. J. 269(1):277-278(1990).  
XX [13]  
RN 11142-12634  
RP MEDLINE; 89345095.  
RA Parquet C., Flouret B., Menguin-Lecreux D., Van Heijenoort J.;  
RT "Nucleotide sequence of the murF gene encoding the UDP-MurNac- pentapeptide  
RL synthetase of *Escherichia coli*";  
RL Nucleic Acids Res. 17:5379-5379(1989).  
XX [14]  
RN 12423-15030  
RP MEDLINE; 90192099.  
RA Ikeda M., Wachi M., Ishino F., Matsuhashi M.;  
RT "Nucleotide sequence involving murD and an open reading frame orf-Y spacing  
RL murF and ftsW in *Escherichia coli*";  
RL Nucleic Acids Res. 18:1058-1058(1990).  
XX [15]  
RN 13392-15020  
RP Menguin-Lecreux D., Van Heijenoort J.;  
RA "Nucleotide sequence of the murD gene encoding the UDP-MurNac- L-Ala-D-gly  
RT synthetase of *Escherichia coli*";  
RL Nucleic Acids Res. 18:183-183(1990).  
XX [16]  
RN 14743-16239  
RP MEDLINE; 90036736.  
RX PUBMED; 2509435.  
RA Ikeda M., Sato T., Wachi M., Jung H.K., Ishino F., Kobayashi Y.,  
RA Matsuhashi M.;  
RT "Structural similarity among *Escherichia coli* FtsW and RodA proteins and  
RL *Bacillus subtilis* SpoVE protein, which function in cell division, cell  
RT elongation, and spore formation, respectively";  
RL J. Bacteriol. 171(11):6375-6378(1989).  
XX [17]  
RN 16094-17806  
RP Menguin-Lecreux D., Texier L., Van Heijenoort J.;  
RA "Nucleotide sequence of the cell-envelope murG gene of *Escherichia coli*";  
RL Nucleic Acids Res. 18:2810-2810(1990).  
XX [18]  
RN 16094-18886  
RP MEDLINE; 90326550.  
RA Ikeda M., Wachi M., Jung H.K., Ishino F., Matsuhashi M.;  
RT "Nucleotide sequence involving murG and murC in the mra gene cluster region  
RL of *Escherichia coli*";  
RL Nucleic Acids Res. 18:4014-4014(1990).  
XX

RN [19] 19464-21952  
RP MEDLINE; 85054557.  
RX PUBMED; 6094474.  
RA Robinson A.C., Kenan D.J., Hatfull G.F., Sullivan N.F., Spiegelberg R.,  
RA Donachie W.D.;  
RT "DNA sequence and transcriptional organization of essential cell division  
RL genes ftsQ and ftsA of *Escherichia coli*: evidence for overlapping  
RL transcriptional units.";  
RL J. Bacteriol. 160(2):546-555(1984).  
XX [20]  
RN 21464-23333  
RP MEDLINE; 86083166.  
RX PUBMED; 3000876.  
RA Yi O.M., Lutkenhaus J.;  
RT "The nucleotide sequence of the essential cell-division gene ftsZ of  
RL *Escherichia coli*";  
RL Gene 36(3):241-247(1985).  
XX [21]  
RN 22964-25011  
RP MEDLINE; 88058745.  
RX PUBMED; 2824434.  
RA Beall B., Lutkenhaus J.;  
RT "Sequence analysis, transcriptional organization, and insertional  
RL mutagenesis of the envA gene of *Escherichia coli*";  
RL J. Bacteriol. 169(12):5408-5415(1987).  
XX [22]  
RN 23989-27799  
RP MEDLINE; 88298644.  
RX PUBMED; 2841285.  
RA Schmidt M., Rollo E., Grodberg I., Oliver D.;  
Alignment Scores:  
Pred. No.: 2,39e-54 Length: 28277  
Score: 986.00 Matches: 230  
Percent Similarity: 63.96% Conservative: 61  
Best Local Similarity: 50.55% Mismatches: 140  
Query Match: 43.63% Indels: 24  
DB: 15 Gaps: 10  
US-09-701-229-2 (1-448) x EC2MIN (1-28277)  
Qy 2 SerLeuileAlaSerAspHisPheArg-----lleValValGlyLeuGlyLysSer 18  
Db 13608 ACGTTAATCATGCTGATTATTCAGGGTAAATAATCTCATATTCGGCCTGGCCTCACC 13667  
Qy 19 GlyMetSerLeuValArgTyrLeuAlaArgGlyLeuProPheAlaValValAspThr 38  
Db 13668 GGGCTTTCCTCGGTGGACTTTTCTCGCTCGCGGTGACGCGCGGTATGATACG 13727  
Qy 39 ArgGluAsnProProGluLeuAlaThrLeuArgAlaGlnTyrProGln---ValGluVal 57  
Db 13728 CGTATGACACCGCTGGCTGATAAATTA-----CCGAAAGCGGTAGAACGC 13775  
Qy 58 ArgCysGlyGluLeuAspAlaGluPheLeuCysSerAlaArgGluLeuTyrValSerPro 77  
Db 13776 CACACGGGCAGTCTGAATGATGATGCTGATGCGGCAGATCTGATGCGCGAGTCCC 13835  
Qy 78 GlyLeuSerLeuArgThrProAlaLeuValGlnAlaAlaAlaLysGlyValArgIleSer 97  
Db 13836 GGTATTCACATGGCGCATCATCTTAAGCGCTCCCGCTGATGCGCGAATCGAATCGTT 13895  
Qy 98 GlyAspIleAspLeuPheAlaArgGluAlaLysAlaProIleValAlaIleThrGlySer 117  
Db 13896 GGGCATATCGAGTGTCTCTCGCGAAACACACACCGCATGTGTGGGATTACCGGTCT 13955  
Qy 118 AsnAlaLysSerThrValThrLeuValGlyGluMetAlaValAlaAlaAspLysArg 137  
Db 13956 AACGGCAAAAGCAGCGGTACACCGCTAGTGGGTGAATGCGGAAAGCGCGGGGTAAAC 14015



6259626	REFERENCE	10	Friedberg, D., Rosenthal, E.R., Jones, J.W. and Calvo, J.M. Characterization of the 3' end of the leucine operon of <i>Salmonella typhimurium</i> Mol. Gen. Genet. 199 (3), 486-494 (1985)
8119399	JOURNAL MEDLINE	20	Cowling, D.W., Bardwell, J.C., Craig, E.A., Woolford, C., Hendrik, R.W. and Gross, C.A. Consensus sequence for <i>Escherichia coli</i> heat shock gene promoters Proc. Natl. Acad. Sci. U.S.A. 82 (9), 2679-2683 (1985)
6262769	REFERENCE	21	Chong, P., Hui, I., Loo, T. and Gilliam, S. Structural analysis of a new GC-specific insertion element IS186 FEBS Lett. 192 (1), 47-52 (1985)
81264207	JOURNAL MEDLINE	22	Birnbaum, M.J., Haspel, H.C. and Rosen, O.M. Cloning and characterization of a cDNA encoding the rat brain glucose-transporter protein Proc. Natl. Acad. Sci. U.S.A. 83 (16), 5784-5788 (1986)
6267039	REFERENCE	23	Sekiguchi, T., Ortega-Cesena, J., Nosoh, Y., Ohashi, S., Tsuda, K. and Kanaya, S. DNA and amino-acid sequences of 3-isopropylmalate dehydrogenase of <i>Bacillus coagulans</i> . Comparison with the enzymes of <i>Saccharomyces cerevisiae</i> and <i>Thermus thermophilus</i> Biochim. Biophys. Acta 867, 36-44 (1986)
6283093	JOURNAL MEDLINE	24	Ohki, M., Tamura, F., Nishimura, S. and Uchida, H. Nucleotide sequence of the <i>Escherichia coli</i> dnaJ gene and purification of the gene product J. Biol. Chem. 261 (4), 1778-1781 (1986)
82216830	REFERENCE	25	Mackie, G.A. Structure of the DNA distal to the gene for ribosomal protein S20 in <i>Escherichia coli</i> K12: presence of a strong terminator and an ISI element Nucleic Acids Res. 14 (17), 6965-6981 (1986)
6283093	JOURNAL MEDLINE	26	Alignment Scores: Pred. NO.: 1.25e-53 Length: 111408 Score: 986.00 Matches: 230 Percent Similarity: 63.96% Conservative: 61 Best Local Similarity: 50.55% Mismatches: 140 Query Match: 43.63% Indels: 24 DB: 1 Gaps: 10
6296778	REFERENCE	27	US-09-701-229-2 (1-448) x ECOL10K (1-111408)
84069770	JOURNAL MEDLINE	28	Qy 2 SerLeuileAlaSerAspHisPheArg -----lleValValGlyLeuGlyLysSer 18
6316258	REFERENCE	29	Db 96739 ACGTAAATCGCTGATTCACGGTAAATAATGTCGATATCGGCTGGCTCACC 96798
6322174	JOURNAL MEDLINE	30	Qy 19 GlyMetSerLeuValArgTyrLeuAlaArgGlyLeuProPheAlaValAlaValAspThr 38
6374664	REFERENCE	31	Db 96799 GGGCTTTCCTCGTGGACTTTTCTCCTCGCTGGTGACGCCGCGGTATGGATACG 96858
6374664	JOURNAL MEDLINE	32	Qy 39 ArgGluAsnProProGluLeuAlaThrLeuArgAlaGlnTyrProGln ---ValGluVal 57
6377309	REFERENCE	33	Db 96859 CGTATGACACCGCTGGCTGGATAAATTA -----CCGAAGCGGTAGAACGC 96906
6377309	JOURNAL MEDLINE	34	Qy 58 ArgCysGlyGluLeuAspAlaGluPheLeuCysSerAlaArgGluLeuTyrValSerPro 77
6377309	REFERENCE	35	Db 96907 CACACGGGACGTCTGAATGATGCTGATGGCGGACAGATCTGATTGTCGCCAGTCC 96966

6259626	REFERENCE	10	Lee, N.L., Gielow, W.O. and Wallace, R.G. Mechanism of araC autoregulation and the domains of two overlapping promoters, Pc and PRAD, in the L-arabinose regulatory region of <i>Escherichia coli</i> Proc. Natl. Acad. Sci. U.S.A. 78 (2), 752-756 (1981)
8119399	JOURNAL MEDLINE	20	Cowling, D.W., Bardwell, J.C., Craig, E.A., Woolford, C., Hendrik, R.W. and Gross, C.A. Consensus sequence for <i>Escherichia coli</i> heat shock gene promoters Proc. Natl. Acad. Sci. U.S.A. 82 (9), 2679-2683 (1985)
6262769	REFERENCE	21	Chong, P., Hui, I., Loo, T. and Gilliam, S. Structural analysis of a new GC-specific insertion element IS186 FEBS Lett. 192 (1), 47-52 (1985)
81264207	JOURNAL MEDLINE	22	Birnbaum, M.J., Haspel, H.C. and Rosen, O.M. Cloning and characterization of a cDNA encoding the rat brain glucose-transporter protein Proc. Natl. Acad. Sci. U.S.A. 83 (16), 5784-5788 (1986)
6267039	REFERENCE	23	Sekiguchi, T., Ortega-Cesena, J., Nosoh, Y., Ohashi, S., Tsuda, K. and Kanaya, S. DNA and amino-acid sequences of 3-isopropylmalate dehydrogenase of <i>Bacillus coagulans</i> . Comparison with the enzymes of <i>Saccharomyces cerevisiae</i> and <i>Thermus thermophilus</i> Biochim. Biophys. Acta 867, 36-44 (1986)
6283093	JOURNAL MEDLINE	24	Ohki, M., Tamura, F., Nishimura, S. and Uchida, H. Nucleotide sequence of the <i>Escherichia coli</i> dnaJ gene and purification of the gene product J. Biol. Chem. 261 (4), 1778-1781 (1986)
82216830	REFERENCE	25	Mackie, G.A. Structure of the DNA distal to the gene for ribosomal protein S20 in <i>Escherichia coli</i> K12: presence of a strong terminator and an ISI element Nucleic Acids Res. 14 (17), 6965-6981 (1986)
6283093	JOURNAL MEDLINE	26	Alignment Scores: Pred. NO.: 1.25e-53 Length: 111408 Score: 986.00 Matches: 230 Percent Similarity: 63.96% Conservative: 61 Best Local Similarity: 50.55% Mismatches: 140 Query Match: 43.63% Indels: 24 DB: 1 Gaps: 10
6296778	REFERENCE	27	US-09-701-229-2 (1-448) x ECOL10K (1-111408)
84069770	JOURNAL MEDLINE	28	Qy 2 SerLeuileAlaSerAspHisPheArg -----lleValValGlyLeuGlyLysSer 18
6316258	REFERENCE	29	Db 96739 ACGTAAATCGCTGATTCACGGTAAATAATGTCGATATCGGCTGGCTCACC 96798
6322174	JOURNAL MEDLINE	30	Qy 19 GlyMetSerLeuValArgTyrLeuAlaArgGlyLeuProPheAlaValAlaValAspThr 38
6374664	REFERENCE	31	Db 96799 GGGCTTTCCTCGTGGACTTTTCTCCTCGCTGGTGACGCCGCGGTATGGATACG 96858
6374664	JOURNAL MEDLINE	32	Qy 39 ArgGluAsnProProGluLeuAlaThrLeuArgAlaGlnTyrProGln ---ValGluVal 57
6377309	REFERENCE	33	Db 96859 CGTATGACACCGCTGGCTGGATAAATTA -----CCGAAGCGGTAGAACGC 96906
6377309	JOURNAL MEDLINE	34	Qy 58 ArgCysGlyGluLeuAspAlaGluPheLeuCysSerAlaArgGluLeuTyrValSerPro 77
6377309	REFERENCE	35	Db 96907 CACACGGGACGTCTGAATGATGCTGATGGCGGACAGATCTGATTGTCGCCAGTCC 96966

```
QY 78 GlyLeuSerLeuArgThrProAlaLeuValGlnAlaAlaAlaLysGlyValArgIleSer 97
Db 96967 GGTATTGCATCGCGCATCCATTAAAGCGCTGCGCGTATGCCGGAATCGAATCGTT 97026

QY 98 GlyAspIleAspLeuPheAlaArgGluAlaLysAlaProIleValAlaIleThrGlySer 117
Db 97027 GCGGATATCGAGTGTCTGTCGCGAAGCAACAGCCGATTTGGCGATTTACCGGTTCT 97086

QY 118 AsnAlaLysSerThrValThrLeuValGlyGluMetAlaValAlaAlaAspLysArg 137
Db 97087 AACGCGAAAGACAGCTCACCGCTAGTGGTGAATGCGGAAGCGCGGGGTTAAC 97146

QY 138 ValAlaValGlyGlyAsnLeuGlyThrProAlaLeuAspLeuAlaAspIleGlu 157
Db 97147 GTTGGTGTGGTGGCAATATGGCTGCGCTGCTGATGCTACTGATGATGATGCTGAA 97206

QY 158 LeuTyrValLeuGluLeuSerPheGlnLeuGluThrCysAspArgLeuAsnAlaGlu 177
Db 97207 CTGTAGCTGCTGGACTGTGAGCTTCCAGCTCGAAGCAACCTCCAGCTTACAGCGGTA 97266

QY 178 ValAlaThrValLeuAsnValSerGluAspHisMetAspArgTyrAsp---GlyMetAla 196
Db 97267 GCAGCGACCATCTGAACGTGACTGAAGATCATATGATCGTATCCGTTGGTTACAA 97326

QY 197 AspTyrHisLeuAlaLysHisArgIlePheArgGlyAlaArgGlnValValValAsnArg 216
Db 97327 CAGTATCTGTCGACGAAAGTGGCATTTACGAAACGCGAAAGTTTGGCTGTTATGCT 97386

QY 217 AlaAlaLeuThrArgProLeuIleAlaAspThrValProCysTrpSerPheGlyLeu 236
Db 97387 GATGATGCTTAAACATGCGATTCGCGTGGGATGAACGCTGGCTGAGCTTGGCGTC 97446

QY 237 AsnLysProAspPheLysAlaPheGlyLeuIleGluLysAspGlyGlnLysTrpLeuAla 256
Db 97447 AACATGGGTGACTATCAC-----CTGAATCATCACAGCGCGAAACCTGGCTGCGG 97497

QY 257 PheGlnPheAspLysLeuLeuProValGlyGluLeuLysIleArgGlyAlaHisAsnTyr 276
Db 97498 GTTAAAGCGGAGAAAGTGTGAATGTGAAGAGATGAACACTTTCGCGGCGACATACTAC 97557

QY 277 SerAsnAlaLeuAlaLeuAlaLeuGlyHisAlaValGlyLeuProPheAspAlaMet 296
Db 97558 ACCATGCGCTGCGCGCTGCGCTGGCAGATGCTGCGAGGTTTACCGGCTGCCAGCAGC 97617

QY 297 LeuGlyAlaLeuLysAlaPheSerGlyLeuAlaHisArgCysGlnTrpValArgGluArg 316
Db 97618 CTGAAGCGTTAAACACATTCATCTGCTGCGCATCGCTTGAAGTTGCTGCGGAGCAT 97677

QY 317 GlnGlyValSerTyrTyrAspSerLysAlaThrAsnValGlyAlaAlaLeuAlaAla 336
Db 97678 AACGCGGTACGTTGATTAACGATTTCGAAGCGCAACGCTGCGCAGTACGGAAGCGGCG 97737

QY 337 IleGluGlyLeuGlyAlaAspIleAspGlyLysLeuValLeuAlaGlyClyAspGly 356
Db 97738 CTGAATGCGCTG-----CACGTACGCGACACTGCGATTTGCTGCGGGCGGCTGCT 97791

QY 357 LysGlyAlaAspPheHisAspLeuArgGluProValAlaArgPheCysArgAla----- 374
Db 97792 AAATCGCGGACTTT-----AGCCCACTGGCGGCTTACCTGAATGGCGATTAAC 97839

QY 375 -----ValValLeuGlyArgAspAlaGlyLeuIleAlaGlnAlaLeuGlyAsnAla 392
Db 97840 GTACGCTGTATGTTTCGGTGGTACGCGCGCGAGCTGGCG-----GCGCTACGCGCGGAA 97896

QY 393 ValProLeuValArgValAlaThrLeuAspGluAlaValArgGlnAlaAlaGluLeuAla 412
Db 97897 GTG---GCAGAACAAACGAACATATGAACAGCGGATGCGCTGCTGCTGCGCGCTGTT 97953

QY 413 ArgGluGlyAspAlaValLeuLeuSerProAlaCysAlaSerLeuAspMetPheLysAsn 432
Db 97954 CAGCGCGGCGATATGGTTCTGCTCTCCCGACGCTGTGCCAGCCTTGATCAGTTCAAGAAC 98013

QY 433 PheGluGluArgGlyArgLeuPheAlaLysAlaValGluGluLeu 447
Db 98014 TTGTAACAACGAGCAATGAGTTTGCCTGCTGCGAAGGAGTTA 98058

RESULT 14
AX189058
LOCUS AX189058
DEFINITION Sequence 259 from Patent WO0148209.
ACCESSION AX189058
VERSION AX189058.1 GI:15142764
KEYWORDS
SOURCE Escherichia coli
ORGANISM Escherichia coli
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; Escherichia.
REFERENCE 1
AUTHORS Forsyth, R.A., Ohlsen, K.L. and Zyskind, J.W.
TITLE Genes identified as required for proliferation of E. coli
JOURNAL Patent: WO 0148209-A 259 05-JUL-2001;
Elitra Pharmaceuticals, Inc. (US)
FEATURES
Location/Qualifiers
1..1317
/organism="Escherichia coli"
/mol_type="genomic DNA"
/db_xref="taxon:562"
1..1317
/note="unnamed protein product"
/codon_start=1
/transl_table=11
/protein_id="CAC50739.1"
/db_xref="GI:15142765"
/translation="MADYQKNNVVIIGLTGLSCVDFFLARGVTPRVMDTRMTPPGL
DKLPAVERHTGSLNDEWLMALADLIVASPIALAHPSLSAAADAGIEIVGDIELFCRE
AQPALVATLGSNGKSTVTTLVGEMAKAVNNGVGNIGLPAIMLLDDSCELYVLELS
SFQLETTSSLOAVATILNVTEDHMDRYFGLQOQRAAKLRIYENAKVCVWADDAIT
MPFADGERCVSFGNMGDYHLNHOQGETWLVKGVKLVNEMKLSGQHNNTNALAA
LALADAAGLPRASSLKALFTTGLPHRFVLEHNGVRWINDSKATNVGSGHTNALNGL
HYDGTLLHLGGDGKSADFSPILRYLNGDNVRLYCFGRDGAQLAALRPEVABQETME
QAMRLAPRVQPDGMVLLSPACASLDQFNFEQGRNEFARLAKELG"
BASE COUNT 294 a 323 c 400 g 300 t
ORIGIN
Alignment Scores:
Pred. No.: 1317
Score: 6.35e-56 Length: 1317
Percent Similarity: 985.50 Matches: 227
Best Local Similarity: 64.33% Conservative: 58
Query Match: 51.24% Mismatches: 137
DB: 43.61% Indels: 21
Gaps: 9
US-09-701-229-2 (1-448) x AX189058 (1-1317)
QY 11 IleValValGlyLeuGlyLysSerGlyMetSerLeuValArgTyrLeuAlaArgGly 30
Db 28 GTCAATATCGCGCTGGCGCTCACCGGGCTTCTCCGTGCGACTTTTCTCCTCGCTCGCGGT 87
QY 31 LeuProPheAlaValAlaAspThrArgGluAsnProGluLeuAlaThrLeuArgAla 50
Db 88 GTGACGCGCGCGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 141
QY 51 GlnTyrProGln---ValGluValArgCysGlyGluLeuAspAlaGluPheLeuCysSer 69
Db 142 -----CCCGAAGCGGTAGAACGCCACACCGCGCTGATGATGATGATGATGATGATGATGAT 195
QY 70 AlaArgGluLeuTyrValSerProGlyLeuSerLeuArgThrProAlaLeuValGlnAla 89
Db 196 GCAGATCTGATGTCGCGCGCTGCGCGCTGCGCGCTGCGCGCTGCGCGCTGCGCGCTGCGCG 255
QY 90 AlaAlaLysGlyValArgIleSerGlyAspIleAspLeuPheAlaArgGluAlaLysAla 109
Db 256 GCTGATGCGCGAATCGAATCGTTGGCGCATATCGAGCTGTTCTGTCGCGAAGCACACGCA 315
QY 110 ProIleValAlaIleThrGlySerAsnAlaLysSerThrValThrLeuValGlyGlu 129
```



```

316 CCGATTGTGGCGATTACCGGTTCTAAACGGCAAAAGACGCGTACACCGCTAGTGGTGAA 375
QY MetAlaValAlaAlaAspLysArgValAlaValGlyGlyAsnLeuGlyThrProAlaLeu 149
Db ATGGCGAAGCGCGGGGTTAACTGGTGGTGGTGGCAATATGGCTCCCTGCGGTG 435
QY AspLeuLeuAlaAspAspIleGluLeuTyrValLeuGluLeuSerSerPheGlnLeuGlu 169
Db ATGCTACTGGATGAGTGTGAACCTGTACGTGGTGAACCTGTGAGCTTCCAGCTGGAA 495
QY ThrCysAspArgLeuAsnAlaGluValAlaThrValLeuAsnValSerGluAspHisMet 189
Db ACCACCTCCAGCTTACAGCGGTAGCAGCGACCATCTCTGAACGTGACTGAAGATCATATG 555
QY AspArgTyrAsp---GlyMetAlaAspTyrHisLeuAlaLysHisArgIlePheArgGly 208
Db GATCGCTATCGTTGGTTTACACAGTATCGTCAGCAAAACCTGCGCATTTACAAAC 615
QY AlaArgGlnValValAlaAsnArgAlaAspAlaLeuThrArgProLeuIleAlaAspThr 228
Db GCGAAAGTTTGGTGGTAAATGCTGATGCTTAAACATGCGGATTCGCGGTGCGGAT 675
QY ValProCysTrpSerPheGlyLeuAsnLysProAspPheLysAlaPheGlyLeuIleGlu 248
Db GAACGCTGCGTCAGCTTGGCGTCAACATGCGGTGACTATCAC-----CTGAATCAT 726
QY GluAspGlyGlnLysTrpLeuAlaPheGlnPheAspLysLeuLeuProValGlyGluLeu 268
Db CAGCAGGCGAAACCTGGCTGCGGTTAAAGCGCGAAGAGTGGTGAATGTGAAGAGATG 786
QY LysIleArgGlyAlaHisAsnTyrSerAsnAlaLeuAlaAlaLeuAlaLeuGlyHisAla 288
Db AAATTTTCCGCGCAGCATACTACACCAATGCGTGGCGGCTGGCGGTGCGCAGATGCT 846
QY ValGlyLeuProPheAspAlaMetIleGlyAlaLeuLysAlaPheSerGlyLeuAlaHis 308
Db GCAGGTTTACCGCTGCGCAGCGCTGAACCGGTAAACCATCTACCTGCTGCTGCGCAT 906
QY ArgCysGlnTrpValArgGluArgGlnGlyValSerTyrTyrAspSerLysAlaThr 328
Db CGCTTGAAGTTGTGCGGACCATACCGCGTACGTGGATTACGATTCAAGTTCGAAGCGACC 966
QY AsnValGlyAlaAlaLeuAlaAlaIleGluGlyLeuGlyAlaAspIleAspGlyLysLeu 348
Db AACCTCGCGCATAGCAGGAGCGCGCTGAATGGCCTG-----CACGTAGACGCGCACACTG 1020
QY ValLeuLeuAlaGlyGlyAspGlyLysGlyAlaAspPheHisAspLeuArgGluProVal 368
Db CATTGTGTGCTGGTGGCGATGGTAAATTCGCGCGACTTT-----AGCCCACTG 1068
QY AlaArgPheCysArgAla-----ValValLeuLeuGlyArgAspAlaGlyLeu 384
Db GCGGTTACCTGAATGCGGATAACGTACGCTGTATTGTTTCGCTGCTGACGCGCGCGAG 1128
QY IleAlaGlnAlaLeuGlyAsnAlaValProLeuValArgValAlaThrLeuAspGluAla 404
Db CTGGCG---GCGCTACGCGCCGGAAGTG---GCAGAACAAACCGAACTATGGAACAGCGC 1182
QY ValArgGlnAlaAlaGluLeuAlaArgGluGlyAspAlaValLeuLeuSerProAlaCys 424
Db ATGCGCTGTGCTGCGGTTCAGCGCGGCGCATATGGTTCGTCTCCCGACCTGT 1242
QY AlaSerLeuAspMetPheLysAsnPheGluGluArgGlyArgLeuPheAlaLysAlaVal 444
Db GCCAGCTTGATCAGTTCAAGAACTTGAACACAGCGCATATGATTTGCCCTCTGGCG 1302
QY GluGluLeu 447
Db AAGAGATTA 1311
RESULT 15
AE005185
```

```

LOCUS AE005185 12791 bp DNA linear BCT 21-MAR-2001
DEFINITION Escherichia coli O157:H7 EDL933 genome, contig 1 of 3, section 9 of 155.
ACCESSION AE005185 AE005174
VERSION AE005185.1 GI:12512782
KEYWORDS
SOURCE
ORGANISM Escherichia coli O157:H7 EDL933
REFERENCE Escherichia coli O157:H7 EDL933
AUTHORS Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Escherichia.
1 (bases 1 to 12791)
Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glasner,J.D., Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A., Posfai,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L., Grobeck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamousis,K., Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C., Welch,R.A. and Blattner,F.R.
Genome sequence of enterohaemorrhagic Escherichia coli O157:H7
Nature 409 (6819), 529-533 (2001)
21074935
11206551
2 (bases 1 to 12791)
Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glasner,J.D., Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A., Posfai,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L., Grobeck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamousis,K., Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C., Welch,R.A. and Blattner,F.R.
Direct Submission
Submitted (22-OCT-2000) Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
Location/Qualifiers
1. 12791
/organism="Escherichia coli O157:H7 EDL933"
/mol_type="genomic DNA"
/strain="EDL933"
/serotype="O157:H7"
/db_xref="taxon:155864"
/note="enterohaemorrhagic"
72. 1076
/gene="fruR"
/note="synonym: Z0090"
72. 1076
/gene="fruR"
/function="regulator; Energy metabolism, carbon Glycolysis"
/note="Residues 1 to 334 of 334 are 100.00 pct identical to residues 1 to 334 of 334 from Escherichia coli K-12 Strain MG1655: B0080"
/codon_start=1
/transl_table=11
/product="transcriptional repressor of fru operon and others"
/protein_id="AAG54384.1"
/db_xref="GI:12512783"
/translation="MKLDEIARLAGSVRTTASYVINGKAKQYRYSKTVKVMVYRE HNYHPNVAAGLRAGTRISIGLVPDLENTYRIANLYERQARGQYQLIACSEDO PDMNRCLEHLLOQVDAIIVSTLPPPHFPYQWANDPFPVIALDRALDREHFTSV GADODAEMLAEELKPEAFTVLYGALPELSVFLREQGFTAKWDKDPREHFTLVAN SYEREAQMLEKWEIHPMQALFTTFSFALLQGMVTLRRDCKLPSDLAIATFGDN ELLDFLQCPVLAAQRHDAERVLEIVLASLDEPRKPKGLTRIKRNLRRGVLSRS "
1678. 2136
/gene="yabB"
/note="synonym: Z0091"
1678. 2136
/gene="yabB"
/function="orf; Unknown function"
/note="Residues 1 to 152 of 152 are 99.34 pct identical to residues 1 to 152 of 152 from Escherichia coli K-12 Strain MG1655: B0081"
/codon_start=1
/transl_table=11
```

```

/feature="orf, hypothetical protein"
/protein_id="AAG54385.1"
/db_xref="GI:12512784"
/translation="MFRGATLVNLSKGRSLVPTRYREOLLENAAGQMVCTIDHHPC
LLLYPLPEWELIPKLSRLSSNNPVRVORLLGHASCOMDGAERLLIAPVLRQHA
GLTKEVWLQGFQNFELWDETTHQQVKEDIDAEQATGOLSERLQDLSL"
2138..3079
/gene="yabc"
/notes="synonym: Z0092"
2138..3079
/gene="yabc"
/function="orf; Not classified"
/notes="Residues 1 to 313 of 313 are 100.00 pct identical
to residues 1 to 313 of 313 from Escherichia coli K-12
Strain MG1655: B0082"
/codon_start=1
/transl_table=11
/product="putative apolipoprotein"
/protein_id="AAG54386.1"
/db_xref="GI:12512785"
/translation="MMENKHTTVDLDEAVNGLINIRPDGIYIDGTFGRGHSRLILSQ
LGEGRLLAIDRDPQAIQIAVAKTIDDPFSIIHGPFSAIGEYVAERDLICKIDGILLDL
GVSSPOLDAAERGFSPMRDGPLMDMDPTRGQSAEWLQTAEDADIWLVKLYGERPF
AKRIARIYERNREQMRTIKELAEVVAATPVKRFKHPATRTFOAVRIWVNSELEE
IQALSSUNLAPGGRSLISIFSHSLERIVKRFMRNSRGQVPGAPLPMTEBQLAKL
GGQLRALGKLPGEAEVNAENPRARSSVLRIARTNA"
3076..3441
/gene="ftsL"
/notes="synonym: Z0093"
3076..3441
/gene="ftsL"
/function="phenotype; Cell division"
/notes="Residues 1 to 121 of 121 are 100.00 pct identical
to residues 1 to 121 of 121 from Escherichia coli K-12
Strain MG1655: B0083"
/codon_start=1
/transl_table=11
/product="cell division protein; ingrowth of wall at
septum"
/protein_id="AAG54387.1"
/db_xref="GI:12512786"
/translation="MISRVTEALS VKVSGMSGHERHALPGVIGDDLLRGGKPLCLFI
CILLTAVTYTAHTRLLTAQREQLVLRDALDIEWRNLIILENALGDHSRVERIAT
EKLQMRVDPQSENIIVVQK"
3457..5223
/gene="ftsI"
/notes="synonym: Z0094"
3457..5223
/gene="ftsI"
/function="enzyme; Cell division"
/notes="Residues 1 to 588 of 588 are 100.00 pct identical
to residues 1 to 588 of 588 from Escherichia coli K-12
Strain MG1655: B0084"
/codon_start=1
/transl_table=11
/product="septum formation; penicillin-binding protein 3;
peptidoglycan synthetase"
/protein_id="AAG54388.1"
/db_xref="GI:12512787"
/translation="MKAATAKTPKROEHEANFISWRPALLCCGICILLALAFILGRVAV
LOVISDPLVKGDMRSLRVQVSTSRGNITRSRPLAVSPVPAIWAIDPREVIDAG
GLISGDRWRKALANALNIPLDQLSARINAPKGRFIYLARQVNPDMADYIKLKLPGIH
LREESRRYPSGVEVTAHLGFTNVDQSQIEGVEKFDKWLTPQGPGRIVRKDRYGRVI
EDISSTDSQAHLALSIDERLQALVIRELNNAFNAESGSAVLVDVNTGEVLAMA
NPSYNPNLSGTPKAEARNRTITDVFEPGSTVKPMVVMALQRVVRENSLVNTIPY
RNGHEIKDVARYSLELTGLVQKSNVGVSKLAMPSSALVDYSRGLCKANLIG
LYGERSGLYPQRSIDIERATFSFGYGLMTPYLQARYATIGISYRPLRISIKVD
PVPGERVFPESIVRTVHMMSVAPLPGGGGVKAALKYRIAKITGTAKKVGPDGRIY
NLIYATAGVAPASQPRFALVVVINDPQAGKYGGAVSAPVFAIGMGVGLRVTMTEPD
ALTTGDKNFEVINQEGTGGRS"
5210..6697
/gene="mure"
/notes="synonym: Z0095"

```

```

CDS
5210..6697
/gene="mure"
/function="enzyme; Cell envelop: Murein sacculus,
peptidoglycan"
/notes="Residues 1 to 495 of 495 are 98.98 pct identical to
residues 1 to 495 of 495 from Escherichia coli K-12 Strain
MG1655: B0085"
/codon_start=1
/transl_table=11
/product="meso-diaminopimelate-adding enzyme"
/protein_id="AAG54389.1"
/db_xref="GI:12512788"
/translation="MADRNLRLDLPVDPAPSRALREMTLDSRVAAAGDLFVAVVGH
QDGRRIYIPQAIQAQVAAITAEAKGEATGEIREMHGVPVIYLSQNERLSALAGREY
HPSDNLXXVGVITGTNGKTTTOLLQAWSOLLGETSVMGTGVLGKVIPTPTTGG
SAVDVOHELAGLVDDQATFCAMEVSSHGLVQHRVAALKFAASVFTNLSDRLDLYHCDM
EYEEAAKWLIIYSEHHCQALINADDEVGRLWLAKLPDVAVSMEDHINPCHGRWLKA
TEVNHDSGATIRFSSWGDESHLGMFAFNVLALLATLALGAGYPLADLLKTTAA
RLQPCGRMEVETAPGKPTVVVDYHTPDALKALQARLHCAGKLCWFCGCGDRDK
GKRLPGAIAEEFADVAVVTDDNPTPEPRAIINDILAGMLDAGHAKVMEDEAEAVTC
AVMOAKENDVVLVAGKGHEDYQIVGNQRLDYSRVTVARLLGVIA"
6694..8052
/gene="mure"
/notes="synonym: Z0096"
6694..8052
/gene="mure"
/function="enzyme; Cell envelop: Murein sacculus,
peptidoglycan"
/notes="Residues 1 to 452 of 452 are 99.33 pct identical to
residues 1 to 452 of 452 from Escherichia coli K-12 Strain
MG1655: B0086"
/codon_start=1
/transl_table=11
/product="D-alanine:D-alanine-adding enzyme"
/protein_id="AAG54390.1"
/db_xref="GI:12512789"
/translation="MISVTLSQTLIDNGELQADITLDAVTTDTTKLTPGCLFVALK
GERFDAHDADQAKAGAGALLVSRPLDIDLQILVKTDLAFGEALAAWVROQVPARY
VALTGSSGKTSVKEMTAAILSQCNLTLYTAGLNNDIGVPMTLRLTPEYDYAVIELG
ANHOGIIVWTVSLTRPEALVNLAAHLEFGSLAGAKAKEIFSGLPENGIAITMN
ADNNNDNQSVIGSRKVRFPSPNAASDFTAIHVTSHGTEFTLQTPGSDVDLLP
LPGRHNIANALAAALSMSVGATLDAIKAGLANLKAIVPGRFLPIQLAENQLLDDSYN
ANGVSTAAVQLVQEMPGIRLVYVGDMAELGAESEACHVQVGVGERAAKIDRVLSVGK
TCSHAISTAVGVEHFAEKTALITRLKLSLAEQQVITILVKGSAAAEVYVRLQENG
TC"
8046..9128
/gene="mray"
/notes="synonym: Z0097"
8046..9128
/gene="mray"
/function="enzyme; Cell envelop: Murein sacculus,
peptidoglycan"
/notes="Residues 1 to 360 of 360 are 99.72 pct identical to
residues 1 to 360 of 360 from Escherichia coli K-12 Strain
MG1655: B0087"

```

## Alignment Scores:

Pred. No.:	Score:	Length:	Matches:
1	1.24e-54	12791	230
2	984.00	230	61
3	63.96%	Conservative:	140
4	50.55%	Mismatches:	24
5	43.54%	Indels:	10
6	1	Gaps:	10

US-09-701-229-2 (1-448) x AE005185 (1-12791)

QY 2 SerLeuIlleAlaSerAspHisPheArg-----IleValValGlyLeuGlyLysSer 18

Db 9122 AGCTTAATCATGCTGCTATACCAGGAGTAAATGTTGTTATTCATTCGCGCCCTGACCCAC 9181

QY 19 GlyMetSerLeuValArgTyrLeuAlaArgGlyLeuProPheAlaValValAspThr 38

Db 9182 GGGCTTTCCTGCTGACTTTTTCCTCGCTCGGGGTGTGACGCCGCCGCTTATGGATACG 9241

```
Qy 39 ArgGluAsnProProGluLeuAlaThrLeuArgAlaGlnTyrProGln---ValGluVal 57
Db 9242 CGTATGACACCGCTGGCTGGATAAATTA-----CCCGAAGCGGTAGAAGC 9289
Qy 58 ArgCysGlyGluLeuAspAlaGluPheLeuCysSerAlaArgGluLeuTyrValSerPro 77
Db 9290 CACACGGCGAGCTGGAATGATGATGGCTGATGGCGCAGATCTGATGGCGAGTCC 9349
Qy 78 GlyLeuSerLeuArgThrProAlaLeuValGlnAlaAlaLysGlyValArgIleSer 97
Db 9350 GGTATTGCTGCGGCATCCATCCTTAAGCGCTCCCGCTGATGCCGGAATCGAATCGTT 9409
Qy 98 GlyAspIleAspLeuPheAlaArgGluAlaLysAlaProIleValAlaIleThrGlySer 117
Db 9410 GGCATATCGAGCTGTCTGTGCGGAAGCAGCAGCAGCAGATTTGTGGGATTACCGTTCT 9469
Qy 118 AsnAlaLysSerThrValThrLeuValGlyGluMetAlaValAlaAlaAspLysArg 137
Db 9470 AACGGCAAAAGCAGCGTCCACCGCTAGTGGTGAATGGCGAAAGCGGGGGTTAAC 9529
Qy 138 ValAlaValGlyGlyAsnLeuGlyThrProAlaLeuAspLeuLeuAlaAspIleGlu 157
Db 9530 GTTGGTGTGGTGCAATATTGGCTGCCGCGTTGATGCTGCTGGATGATGAGTGGGAA 9589
Qy 158 LeuTyrValLeuGluLeuSerSerPheGlnLeuGluThrCysAspArgLeuAsnAlaGlu 177
Db 9590 CTGTACGTACTGGAACTGTCCAGCTTCCAGCTGCGAAACCCAGCTTACAGCGGTA 9649
Qy 178 ValAlaThrValLeuAsnValSerGluAspHisMetAspArgTyrAsp---GlyMetAla 196
Db 9650 GCACGACCATCTCGACGTGACTGAAGATCATATGATGCTATCGCTTGGTTTACAG 9709
Qy 197 AspTyrHisLeuAlaLysHisArgIlePheArgGlyAlaArgGlnValValAsnArg 216
Db 9710 CAATACCGTGGCGCAAAATTCGTATTACGAAATGCGAAAGTTTGTGCTCAATGCT 9769
Qy 217 AlaAspAlaLeuThrArgProLeuIleAlaAspThrValProCysTyrSerPheGlyLeu 236
Db 9770 GATGATGCTTTAACAATGCCGATTCGCGGTGCGGATGAACGCTGCGTCAGCTTTGGCGTC 9829
Qy 237 AsnLysProAspPheLysAlaPheGlyLeuIleGluAspGlyGlnLysTyrLeuAla 256
Db 9830 AACATGGGTGACTATCAC-----CTGNATCATCAGCAGCGCGGAAACCTGGCTGGG 9880
Qy 257 PheGlnPheAspLysLeuLeuProValGlyGluLeuLysIleArgGlyAlaHisAsnTyr 276
Db 9881 GTGAAAGCGGAGAAAGTGTCAACGTGAAGAAATGAACCTTCCGGGACAGCAATACTAC 9940
Qy 277 SerAsnAlaLeuAlaAlaLeuAlaLeuGlyHisAlaValGlyLeuProPheAspAlaMet 296
Db 9941 ACCAATGCGTGGCGCGCTGGCGCTGCGAGATCCCGCAGGTTTACCGCGCTGCCAGCAGC 10000
Qy 297 LeuGlyAlaLeuLysAlaPheSerGlyLeuAlaHisArgCysGlnTyrValArgGluArg 316
Db 10001 CTGAACGGTTAACCATTCATCTGCTGGCTGCGGATCGCTTTGAAGTTGTGCTGGAGCAT 10060
Qy 317 GlnGlyValSerTyrTyrAspSerLysAlaThrAsnValValGlyAlaAlaLeuAlaAla 336
Db 10061 AACGGCTACGTTGATTAAGATTTCGAACGACCAACCGTCGCGCAGTACGGAAGCGCG 10120
Qy 337 IleGluGlyLeuGlyAlaAspIleAspGlyLysLeuValLeuLeuAlaGlyGlyAspGly 356
Db 10121 CTGAATGGCTG-----CAGTAGACGGCAGCGTGCATTTGTGCTGGCGGGGATGGT 10174
Qy 357 LysGlyAlaAspPheHisAspLeuArgGluProValAlaArgPheCysArgAla----- 374
Db 10175 AAATCGCGGACTTT-----AGCCACTGGCGGCTTACCTCAATGGCGATAAAC 10222
Qy 375 -----ValValLeuLeuGlyArgAspAlaGlyLeuIleAlaGlnAlaLeuGlyAsnAla 392
Db 10223 GTACGTCGTGATTGTTGCGTCTGTGAC---GGCGTGAGCTGGCGGCGCTACGCCCGGAA 10279
Qy 393 ValProLeuValArgValAlaThrLeuAspGluAlaValArgGlnAlaAlaGluLeuAla 412
```

```
Db 10280 GTG---GCAGAACAAACCGAAACTATGGAACAGCGATGCGCTTGCTCCGCGTGT 10336
Qy 413 ArgGluGlyAspAlaValLeuLeuSerProAlaCysAlaSerLeuAspMetPheLysAsn 432
Db 10337 CAGCCGGCGGATATGGTTCTACTCTCCCGGCGCTGTGCCAGCCTTGATCAGTTCAAGAAC 10396
Qy 433 PheGluGluArgGlyArgLeuPheAlaLysAlaValGluGluLeu 447
Db 10397 TTTGAACAACGAGCAATGAGTTTGCCCGTCTGCGCGAAGGAGTTA 10441
```

Search completed: August 14, 2003, 08:41:41  
Job time : 4825 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 14, 2003, 03:24:05 ; Search time 5305 Seconds  
(without alignments)  
11181.708 Million cell updates/sec

Title: US-09-701-229-1  
Perfect score: 1450  
Sequence: 1 cgtgctgacggcctcgcca.....tgttgagcggcagcgcatc 1450

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 2888711 seqs, 20454813386 residues

Word size : 0  
Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : GenEmbl.\*

- 1: gb\_ba.\*
- 2: gb\_htg.\*
- 3: gb\_in.\*
- 4: gb\_om.\*
- 5: gb\_ov.\*
- 6: gb\_pat.\*
- 7: gb\_ph.\*
- 8: gb\_pl.\*
- 9: gb\_pr.\*
- 10: gb\_ro.\*
- 11: gb\_sts.\*
- 12: gb\_sy.\*
- 13: gb\_un.\*
- 14: gb\_vl.\*
- 15: em\_ba.\*
- 16: em\_fun.\*
- 17: em\_hum.\*
- 18: em\_in.\*
- 19: em\_mu.\*
- 20: em\_om.\*
- 21: em\_or.\*
- 22: em\_ov.\*
- 23: em\_pat.\*
- 24: em\_ph.\*
- 25: em\_pl.\*
- 26: em\_ro.\*
- 27: em\_sts.\*
- 28: em\_un.\*
- 29: em\_vl.\*
- 30: em\_htg\_hum.\*
- 31: em\_htg\_inv.\*
- 32: em\_htg\_other.\*
- 33: em\_htg\_mus.\*
- 34: em\_htg\_pln.\*
- 35: em\_htg\_rtd.\*
- 36: em\_htg\_mam.\*
- 37: em\_htg\_vrt.\*
- 38: em\_sy.\*
- 39: em\_htgo\_hum.\*
- 40: em\_htgo\_mus.\*
- 41: em\_htgo\_other.\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
C 1	1449	99.9	24000	1	AE004856	AE004856 Pseudomon
C 2	1248	86.1	5280	1	AY008276	AY008276 Pseudomon
C 3	44	3.0	301995	1	AY016779	AY016779 Sequence
C 4	24	1.7	1317	6	AX189058	AX189058 Escherichia
C 5	24	1.7	1629	1	ECMURD	X17609 Escherichia
C 6	24	1.7	2608	1	ECMUROY	X51584 Escherichia
C 7	24	1.7	11441	1	AE015046	AE015046 Shigella
C 8	24	1.7	12791	1	AE005185	AE005185 Escherich
C 9	24	1.7	21757	1	AE000118	AE000118 Escherich
C 10	24	1.7	28277	6	AX191720	AX191720 Sequence
C 11	24	1.7	28277	15	EC2MIN	X55034 E. coli 2 m
C 12	24	1.7	111408	1	EC0110K	D10483 Escherichia
C 13	24	1.7	248308	2	AC073937	AC073937 Mus muscu
C 14	24	1.7	259498	2	AC020876	AC020876 Mus muscu
C 15	24	1.7	281530	1	AP002350	AP002350 Escherich
C 16	24	1.7	290029	1	AE016978	AE016978 Shigella
C 17	24	1.7	300409	1	AE016755	AE016755 Escherich
C 18	24	1.7	329709	1	AP002997	AP002997 Mesorhizo
C 19	23	1.6	713	9	HSA328412	AJ28412 Homo sapi
C 20	23	1.6	8599	1	AB004563	AB004563 Terrabact
C 21	23	1.6	110000	2	AC141230_2	Continuation (3 of
C 22	23	1.6	311600	1	AE016871	AE016871 Pseudomon
C 23	23	1.5	31495	1	BJU33883	U33883 Bradyrhizob
C 24	22	1.5	207050	1	AL646063	AL646063 Ralstonia
C 25	22	1.5	300900	1	AP005939	AP005939 Bradyrhiz
C 26	21	1.4	9077	1	AB032799	AB032799 Chromobac
C 27	21	1.4	10094	1	AF172851	AF172851 Chromobac
C 28	21	1.4	11985	1	PPY19106	Y19106 Pseudomonas
C 29	21	1.4	13376	1	U72354	U72354 Pseudomonas
C 30	21	1.4	69301	1	SCU82965	SCU82965 Streptococ
C 31	21	1.4	86896	1	RCU57682	RCU57682 Rhodobacte
C 32	21	1.4	132469	2	AP003926	AP003926 Oryza sat
C 33	21	1.4	152460	8	AP004617	AP004617 Oryza sat
C 34	21	1.4	186217	10	AC122310	AC122310 Mus muscu
C 35	21	1.4	310550	1	SC0939113	SC0939113 Streptoc
C 36	20	1.4	2076	6	AX196006	AX196006 Sequence
C 37	20	1.4	3511	9	HSSKIR	X15218 Human ski o
C 38	20	1.4	3772	1	MELOSVM12	AF049243 Mesorhizo
C 39	20	1.4	4583	1	SMASFOABC	M33815 S.marcescen
C 40	20	1.4	7313	1	AF183959	AF183959 Pseudomon
C 41	20	1.4	10029	1	AE012267	AE012267 Xanthomon
C 42	20	1.4	10684	1	AE009365	AE009365 Agrobacte
C 43	20	1.4	10801	1	AE012280	AE012280 Xanthomon
C 44	20	1.4	12036	1	AE008249	AE008249 Agrobacte
C 45	20	1.4	12296	1	AF031417	AF031417 Pseudomon

ALIGNMENTS

RESULT 1  
AE004856/c  
LOCUS  
DEFINITION Pseudomonas aeruginosa PAO1, section 417 of the complete genome.  
ACCESSION AE004856 AE004091  
VERSION AE004856.1 GI:9950633  
KEYWORDS Pseudomonas aeruginosa PAO1  
SOURCE Pseudomonas aeruginosa PAO1  
ORGANISM Bacteria: Proteobacteria: Gammaproteobacteria; Pseudomonadales;  
Pseudomonadaceae; Pseudomonas.  
REFERENCE 1 (bases 1 to 24000)  
AUTHORS Stover,C.K., Pham,X.-Q.T., Erwin,A.L., Mizoguchi,S.D., Warrenner,P., Hickey,M.J., Brinkman,F.S.L., Hufnagle,W.O., Kowalik,D.J.,

**TITLE**  
JOURNAL  
MEDLINE  
PUBMED  
REFERENCE  
AUTHORS

Lagrou, M., Garber, R.L., Goltry, L., Tolentino, E., Westbrock-Wadman, S., Yuan, Y., Brody, L.L., Coulter, S.N., Folger, K.R., Kas, A., Larbig, K., Lim, R.M., Smith, K.A., Spencer, D.H., Wong, G.K.-S., Wu, Z., and Paulsen, I.T.  
Complete genome sequence of *Pseudomonas aeruginosa* PA01, an opportunistic pathogen  
Nature 406 (6799), 599-964 (2000)  
20437337  
10984043  
2 (bases 1 to 24000)  
Stover, C.K., Pham, X.-Q.T., Erwin, A.L., Mizoguchi, S.D., Warren, P., Hickey, M.J., Brinkman, F.S.L., Huftnagle, W.O., Kowalik, D.J., Lagrou, M., Garber, R.L., Goltry, L., Tolentino, E., Westbrock-Wadman, S., Yuan, Y., Brody, L.L., Coulter, S.N., Folger, K.R., Kas, A., Larbig, K., Lim, R.M., Smith, K.A., Spencer, D.H., Wong, G.K.-S., Wu, Z., Paulsen, I.T., Reizer, J., Saiter, M.H., Hancock, R.E.W., Lory, S. and Olson, M.V.  
Direct Submission  
Submitted (16-MAY-2000) Department of Medicine and Genetics, University of Washington Genome Center, University Of Washington, Box 352145, Seattle, WA 98195, USA  
3 (bases 1 to 24000)

**TITLE**  
JOURNAL  
MEDLINE  
PUBMED  
REFERENCE  
AUTHORS  
CONSTRM  
TITLE  
JOURNAL  
COMMENT

*Pseudomonas aeruginosa* Community Annotation Project (PseudocAP)  
Direct Submission  
Submitted (04-FEB-2003) Department of Molecular Biology and Biochemistry, Simon Fraser University, 8888 University Dr., Burnaby, British Columbia V5A 1S6, Canada  
-----  
This represents the February 3, 2003 version of the continually updated, reviewed, *Pseudomonas aeruginosa* PA01 genome annotation, from PseudocAP (see <http://www.pseudomonas.com> for latest updates and links to alternate annotations). PseudocAP is coordinated by Fiona S.L. Brinkman (Simon Fraser University, Canada) and Robert E.W. Hancock (University of British Columbia, Canada). We welcome submission through [www.pseudomonas.com](http://www.pseudomonas.com) of any proposed changes.

'Protein name confidence' is used to rate our confidence of the accuracy of the protein name.  
Class 1: Function experimentally demonstrated in *P. aeruginosa*.  
Class 2: Function of highly similar gene experimentally demonstrated in another organism (and gene context consistent in terms of pathways its involved in, if known).  
Class 3: Function proposed based on presence of conserved amino acid motif, structural feature or limited sequence similarity to an experimentally studied gene.  
Class 4: Homologs of previously reported genes of unknown function, or no similarity to any previously reported sequences.

**FEATURES**  
source  
Location/Qualifiers  
1. .24000  
/organism="Pseudomonas aeruginosa PA01"  
/mol\_type="genomic DNA"  
/strain="PA01"  
/db\_xref="taxon:208964"  
complement(91. .1308)  
/gene="argJ"  
/locus\_tag="PA4402"  
complement(91. .1308)  
/gene="argJ"  
/locus\_tag="PA4402"  
/note="Protein name confidence: Class 2 (high similarity to functionally studied protein); Subcellular localization: Cytoplasmic; Subcellular localization confidence: Class 2; Alt protein name: ornithine transacetylase"  
/codon\_start=1  
/transl\_table=11  
/product="glutamate N-acetyltransferase"  
/protein\_id="AAG07790.1"  
/db\_xref="GI:9950634"  
/translation="MAVGLPLSTLHPVPGFELGIASGKIPGRKDVVVMRCAGSS VAGVFTLNAFCAPVTLAKQRFELGEVRYLLTWTGNAGTGEAGLAAATCAKLAEL

**gene**  
complement(1455. .4205)  
/gene="secA"  
/locus\_tag="PA4403"  
complement(1455. .4205)  
/gene="secA"  
/locus\_tag="PA4403"  
/note="Protein name confidence: Class 2 (high similarity to functionally studied protein)"  
/codon\_start=1  
/transl\_table=11  
/product="secretion protein SecA"  
/protein\_id="AAG07791.1"  
/db\_xref="GI:9950635"  
/translation="MFAPILLKFLGSKNERDVKRAKAVQAINALEPQMVALSDEQLK AKTAEQYQYAKGETLDQLPEAFVAVVYVNDYLARRDNMRPLVGLSGVVTIPF RTGEGTLVGLTVLNLALSGKGVHVYVNDYLARRDNMRPLVGLSGVVTIPF QPPEDKRAAYAADITGTNNEFGDYLRDNMAFSLDDKFORNELFAVDEVDLIDE ARTPLISGAEDSSELIYKINKALIPLAKQVVEEGKPTGEEGHYSIDERTVELNE QHQFIEDLLSQNLGEGESLYSAHNLSLLHYAALRAHTLHFRVNEYVIGDQIL LIDHTGRTPMGRRLSEGHLQAEKEGLPIQAESQTLASTTQNTFRLYNKLGKMTG TADTEAFRQIYGLDVVITPHTPARKDFNDLVLTQEEKYAAITDITKQCAQLR PILVGSATSESVESVKLQEGIEHKVLNAKYEKEAIIAQAGAPSVTIATNMAG RGTDIILGNGVEVALENPTEQIOAIKAQWKHQVQVIEAGGLHVIASERHESR IDNQLGRAGQDPCSSRYLSLEDNLRIFASDRVKNPKMKGALQSGGSAIEHRAVT LATEKAKRQVEGRNFDIKQLLEDVANEQRKVIYHMRNLISAEDVGTEIREFEE TSATINQHIPQSLPEQDIELEAALYSDFAVRPIQQMLDEDKLYBETIRLEE EQYVAAYEKEELAGAEALRAFEKMLRLVLDLWDKHLSTMDHLRHGHLRGYAKNR PKQYKRESFTLFQELDSIKRDTIRVLSHVQVRREDPAEEARLRRELAARAKQF QHAEAPSMQAVAGEEELPEGPAPVVPLEPVRNEQKIGRNEPCPGSGKKYKHCQ LD"

**gene**  
complement(4339. .5262)  
/locus\_tag="PA4404"  
complement(4339. .5262)  
/locus\_tag="PA4404"  
/note="Protein name confidence: Class 4 (homologs of previously reported genes of unknown function, or no similarity to any previously reported sequences)"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="AAG07792.1"  
/db\_xref="GI:9950636"  
/translation="MHIIFLSRHHGAARSVHLQPRLLMLAGACVVGASLLAGFMGNW LFRAPAPAGIEVDAAALQAQSAEISSAREDAQMDALSVHLGELQARLVRLDALGER LTELADVDPEDFSLVSGGGYDEPLEGSAYAPPSPMATLDELAARLDTREQOLEYL EQLLADKHISQSTYLQAGQPVQGYISSPGRVNVPSGRLSMHKGIIDFAPAGSDVYA VAAQVVTWGRKNGYGVVEVGHADGYTTLIAHNQKNLVEVDLVKRGQVLAKVGSYG RSTGYHVHFEVMDGRVMNPQSFARIASVSE"

**gene**  
complement(5409. .5804)  
/locus\_tag="PA4405"  
complement(5409. .5804)  
/locus\_tag="PA4405"  
/note="Protein name confidence: Class 4 (homologs of previously reported genes of unknown function, or no similarity to any previously reported sequences)"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="AAG07793.1"  
/db\_xref="GI:9950637"  
/translation="MLFAEQRICRLQQLLESLOQAPAPYCHVSWRGCLLIIVTD GHWATRLYQORRLQQLQAEFEFLAKILFKVQPTTSTQRAARSLTUSAVALNL RESAEIVGNPRLEALERLASRAEKSRQE"

**gene**  
complement(5865. .6776)  
/gene="lpxC"  
/locus\_tag="PA4406"  
/note="synonyms: envA, asmB"

CDS	complement(5865. .6776) /gene="lpxC" /locus_tag="PA4406" /note="Protein name confidence: Class 1 (function experimentally demonstrated in P. aeruginosa); Alt protein name: EnvA protein" /codon_start=1 /transl_table=1 /product="UDP-3-O-acyl-N-acetylglucosamine deacetylase" /protein_id="AA070794.1" /db_xref="GI:9950638" /translation="MIKQRTILNIIRATGVGLSHGKVIYTLKPAPVDVTGIVFCRTDL DPVVEIPARENVGETHTMSTLTKVGDVKTVEHLISAMAGLIDINAYVELSASEVPI MDGSAFVFLIQSAGEQEAARKFIRIKREVSVEGDKRAVEVPFDGKVFSEIYD DHPVFRTOQASVDFSTSVFVRSARTFGPMRDTEYLRSONLALGGSVKNIAVVD ENRVLNEGLRYEDEFVKHILDAIGDLYLLGNSLIGEFRGFSGHALNQQLRLTLLA DKDAWEVTFEDARTAPISYMRPAAV"
gene	complement(6889. .8073) /gene="ftsZ" /locus_tag="PA4407" /locus_tag="PA4407" /gene="ftsZ" /locus_tag="PA4407" /note="Protein name confidence: Class 2 (high similarity to functionally studied protein)" /codon_start=1 /transl_table=1 /product="cell division protein FtsZ" /protein_id="AA070795.1" /db_xref="GI:9950639" /translation="MFELDVNIATVAVIKVIGVGGGNAVNHMAKNVGVFEICAN TDAQALNIARTVLIQPGVTGKLGAGANPEVQAALEDREIRISVLEGADVFIT TGMGGTGTGAAPITIAVAKEMGILTVAVTRPPFEGEKRMOIADGIRALAEVDS LITIPNEKLITILGKDSLAAAPAKADVDLAGAVRGISDIIKRPGLNVDFADKVTVM SEMGMAMGTGCASEPNAREATEAARNPLEDVNLQAGRGILVNTAGPDLISLGEY SDVGNIIEQFASERATVKVGTVIDADRDELHVIYVAGLGARLEKPKVYVDNIVQGS AD"
gene	complement(8124. .9377) /gene="ftsA" /locus_tag="PA4408"
CDS	complement(6889. .8073) /locus_tag="PA4407" /locus_tag="PA4407" /gene="ftsZ" /locus_tag="PA4407" /note="Protein name confidence: Class 2 (high similarity to functionally studied protein)" /codon_start=1 /transl_table=1 /product="cell division protein FtsZ" /protein_id="AA070795.1" /db_xref="GI:9950639" /translation="MFELDVNIATVAVIKVIGVGGGNAVNHMAKNVGVFEICAN TDAQALNIARTVLIQPGVTGKLGAGANPEVQAALEDREIRISVLEGADVFIT TGMGGTGTGAAPITIAVAKEMGILTVAVTRPPFEGEKRMOIADGIRALAEVDS LITIPNEKLITILGKDSLAAAPAKADVDLAGAVRGISDIIKRPGLNVDFADKVTVM SEMGMAMGTGCASEPNAREATEAARNPLEDVNLQAGRGILVNTAGPDLISLGEY SDVGNIIEQFASERATVKVGTVIDADRDELHVIYVAGLGARLEKPKVYVDNIVQGS AD"
gene	complement(8124. .9377) /gene="ftsA" /locus_tag="PA4408"
Query Match	99.9%; Score 1449; DB 1; Length 24000;
Best Local Similarity	100.0%; Pred. No. 0;
Matches 1449; Conservative	0; Mismatches 0; Indels 0; Gaps 0;
Qy	2 GTGCTGATGGCCCTGGCACCTTGAAGCTGGTGGAGSACGAAGAGAGCATGAGCCTCAT 61
Db	16314 GTGCTGATGGCCCTGGCACCTTGAAGCTGGTGGAGSACGAAGAGAGCATGAGCCTCAT 16255
Qy	62 CGCCTCCGACCATTCGCGCATCGTGTGGCGCTCGGCAAGAGCGGCATGTCCTTGGTGGC 121
Db	16254 CGCCTCCGACCATTCGCGCATCGTGTGGCGCTCGGCAAGAGCGGCATGTCCTTGGTGGC 16195
Qy	122 CTACTTGGCGCGCGCGGCTTGCCTTTCGCGGTGGTGCATACCGAGAGAACCCGCGGA 181
Db	16194 CTACTTGGCGCGCGCGGCTTGCCTTTCGCGGTGGTGCATACCGAGAGAACCCGCGGA 16135
Qy	182 GCTGGCCACCTCGTGGCCAGTATCCGAGGTGGAAGTGGTGGCGGCAACTCGACGC 241
Db	16134 GCTGGCCACCTCGTGGCCAGTATCCGAGGTGGAAGTGGTGGCGGCAACTCGACGC 16075
Qy	242 CGAGTTCCTTCGCTCCGCGCCGCAACTCTATGTGAGCCCGCGGCTTGTGCTGGCGACCCC 301
Db	16074 CGAGTTCCTTCGCTCCGCGCCGCAACTCTATGTGAGCCCGCGGCTTGTGCTGGCGACCCC 16015
Qy	302 TGCCTGTGTACAGCGCGCGCGGAAAGCGGTTCGCGATCTCCGGTGACATCTCTTCGC 361
Db	16014 TGCCTGTGTACAGCGCGCGCGGAAAGCGGTTCGCGATCTCCGGTGACATCTCTTCGC 15955
Qy	362 CCGCGAGCGGAGCGCGCGCGATCGTCGCATCACCGGTTCGAACGCGAGAGCACCGTGAC 421
Db	15954 CCGCGAGCGGAGCGCGCGCGATCGTCGCATCACCGGTTCGAACGCGAGAGCACCGTGAC 15895

Qy	422 CACCTTGGTGGCGAAATGGCGTGGCGGCAAGCGTGTGCCGTGGCGGCAACCT 481
Db	15894 CACCTTGGTGGCGAAATGGCGTGGCGGCAAGCGTGTGCCGTGGCGGCAACCT 15835
Qy	482 CGGCACCCCGCGCTCGCACTGTGGCGGAGACATCGAGCTGTACGTGTGGAGCTGTC 541
Db	15834 CGGCACCCCGCGCTCGCACTGTGGCGGAGACATCGAGCTGTACGTGTGGAGCTGTC 15775
Qy	542 GAGCTTCAGCTGGAACCTCGCATCGCTCAAGCGGAGGTGGGACCGTCTGAACGT 601
Db	15774 GAGCTTCAGCTGGAACCTCGCATCGCTCAAGCGGAGGTGGGACCGTCTGAACGT 15715
Qy	602 CAGCGAAGACCATATGATCGCTACGAGCGCATGGCTGACTACCACTTGGCCAAAGCACCG 661
Db	15714 CAGCGAAGACCATATGATCGCTACGAGCGCATGGCTGACTACCACTTGGCCAAAGCACCG 15655
Qy	662 GATCTTCGGGTGCCGCCAGGTGCTGTGTAATCGGCGCCGATGCCCTGACCCGACCGCT 721
Db	15654 GATCTTCGGGTGCCGCCAGGTGCTGTGTAATCGGCGCCGATGCCCTGACCCGACCGCT 15595
Qy	722 GATCGCGATACCGTGGCGTCTGCTGGCTTGAACGCGGACTTCAAGGCTTT 781
Db	15594 GATCGCGATACCGTGGCGTCTGCTGGCTTGAACGCGGACTTCAAGGCTTT 15535
Qy	782 CGGCTTGTATCGAGGAAGACGCCAGAGTGGCTGGCTTCCAGTTCGACAAAGCTGCTGCC 841
Db	15534 CGGCTTGTATCGAGGAAGACGCCAGAGTGGCTGGCTTCCAGTTCGACAAAGCTGCTGCC 15475
Qy	842 GTTTGGGCACTGAAGATCCGTGGCGCCACAATATTTCCAAAGGCTTGGCGGCTGGC 901
Db	15474 GTTTGGGCACTGAAGATCCGTGGCGCCACAATATTTCCAAAGGCTTGGCGGCTGGC 15415
Qy	902 GCTGGGCACTGGCTGGCGTCCGCTTCGAGCGCATGCTCGCGCTGAGAGGCTTTC 961
Db	15414 GCTGGGCACTGGCTGGCGTCCGCTTCGAGCGCATGCTCGCGCTGAGAGGCTTTC 15355
Qy	962 CGGCTTGGCTCATCGCTGCCAGTGGTACGCGAGCGGCGGCTGAGCTACTACGACGA 1021
Db	15354 CGGCTTGGCTCATCGCTGCCAGTGGTACGCGAGCGGCGGCTGAGCTACTACGACGA 15295
Qy	1022 TTCAAAGGCCAACCAAGCTGGCGCGCGCTGGCGGCGATCGAGGGGCTGGGTGGCGACAT 1081
Db	15294 TTCAAAGGCCAACCAAGCTGGCGCGCGCTGGCGGCGATCGAGGGGCTGGGTGGCGACAT 15235
Qy	1082 CGAGGCAAGCTGGTGTGCTCGCGCGGAGAGCGCAAGGCGCGCATTTCCATGACCT 1141
Db	15234 CGAGGCAAGCTGGTGTGCTCGCGCGGAGAGCGCAAGGCGCGCATTTCCATGACCT 15175
Qy	1142 CGCGAGCGGCTCGCGCTTCTCGCGGCGGTGGTACTTGGCCGTGACGCCGGCT 1201
Db	15174 CGCGAGCGGCTCGCGCTTCTCGCGGCGGTGGTACTTGGCCGTGACGCCGGCT 15115
Qy	1202 GATTGCCAGGCACTGGGCAAGCGGTACCGCTGGTGGCGCTGCGAAGCTGGAGCAAGC 1261
Db	15114 GATTGCCAGGCACTGGGCAAGCGGTACCGCTGGTGGCGCTGCGAAGCTGGAGCAAGC 15055
Qy	1262 AGTCCGCGAGCGCGCGAGCTGGCGGAAAGGAGTGCAGTGTGTTGTGCGCGGCTG 1321
Db	15054 AGTCCGCGAGCGCGCGAGCTGGCGGAAAGGAGTGCAGTGTGTTGTGCGCGGCTG 14995
Qy	1322 CGCGAGCGCTGACATGTTTCAAGAACTTCGAAGAACCGCGAGCGCTTGTTCGCAAGCGCT 1381
Db	14994 CGCGAGCGCTGACATGTTTCAAGAACTTCGAAGAACCGCGAGCGCTTGTTCGCAAGCGCT 14935
Qy	1382 AGAGAGCTACGCTGATGCTGCTGGTGTTCGCCCTTCCCGCTGCGCGCTGTGAGCGCG 1441
Db	14934 AGAGAGCTACGCTGATGCTGCTGGTGTTCGCCCTTCCCGCTGCGCGCTGTGAGCGCG 14875
Qy	1442 CACGGCATC 1450
Db	14874 CACGGCATC 14866

```

RESULT 2
AY008276
LOCUS      5280 bp      DNA      linear      BCT 19-JUL-2001
DEFINITION Pseudomonas aeruginosa MurE (murE), MurF (murF), MurY (murY), and
            MurD (murD) genes, complete cds.
ACCESSION AY008276
VERSION   AY008276.2 GI:12007715
KEYWORDS
SOURCE    Pseudomonas aeruginosa
ORGANISM  Pseudomonas aeruginosa
            Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
            Pseudomonadaceae; Pseudomonas.
REFERENCE 1 (bases 1 to 5280)
            Azzolina,B.A., Yuan,X., Anderson,M.S. and El-Sherbeini,M.
            The cell wall and cell division gene cluster in the Mra operon of
            Pseudomonas aeruginosa: cloning, production, and purification of
            active enzymes
            Protein Expr. Purif. 21 (3), 393-400 (2001)
            21178826
            11281713
REFERENCE 2 (bases 1 to 5280)
            El-Sherbeini,M. and Azzolina,B.
            Pseudomonas aeruginosa murY, murD, murE, and murF genes
            Unpublished
            3 (bases 1 to 5280)
            El-Sherbeini,M. and Azzolina,B.
            Direct Submission
            Submitted (20-SEP-2000) Biochemistry, Merck Research Laboratories,
            126 E. Lincoln Avenue, Rahway, NJ 07065, USA
            4 (bases 1 to 5280)
            El-Sherbeini,M. and Azzolina,B.
            Direct Submission
            Submitted (02-JAN-2001) Biochemistry, Merck Research Laboratories,
            126 E. Lincoln Avenue, Rahway, NJ 07065, USA
            Sequence update by submitter
            On Jan 2, 2001 this sequence version replaced gi:10719753.
FEATURES
            source
            1..5280
            /organism="Pseudomonas aeruginosa"
            /mol_type="genomic DNA"
            /db_xref="taxon:287"
            1..1464
            /gene="murE"
            /gene="murF"
            1..1464
            /notes="UDP-N-acetylmuramoyl-L-alanyl-D-glutamate.meso-2,
            6-diaminopimelate ligase"
            /codon_start=1
            /product="MurE"
            /protein_id="AAG45235.1"
            /db_xref="GI:12007716"
            /translation="WPMSLSQLFPQAEERDLRLTLDHSGVRPVDLFTVPGGHQDG
            RHADALTKGATAVAEAGAGELPPSDAPLIAVKGAAQLASAVAGRYGEPSPRGLD
            LGVGTNGKTSVQLAQALDLGERCGIVTGLTGFYGALESGRHTTDPPLAVQAT
            LATLKQAGARAVMEVSSGLDQGRVAALGFDAIVFTNLSDRLDHYHGSMEAYAAKA
            KLFAPDLRCRVINLDDFRRLAGEODSELITYSLTDSAPLYCREASPTASPTAR
            RSSIAKOCILRSPILGRNLNLAAGCALLGLGYPGLDILRTLPQLGPGVRQRL
            GGGGPLVYVDVAHPFDALKEKLEALRPHAAARLCLFCGCGDGRDGRPLMAAIAER
            LADEVLTQDNPTEASAAITADIRKFAAADKVTFPLPSRGEAIAHLIASAAVDDVVL
            LAGKHEDYQEDGVRNHPSPDIEQAERLAAWEPVHA"
            1457..2833
            /gene="murF"
            1457..2833
            /gene="murF"
            /gene="murF"
            /note="recombinant enzyme catalyzed the ATP-dependent
            addition of D-alanine-D-alanine to the
            UDP-N-acetylmuramyl-L-alanine-D-glutamate-m-Dap precursor"
            /codon_start=1
            /transl_table=11
            /product="MurF"
            /protein_id="AAG45236.1"

```

```

/db_xref="GI:12007717"
/translation="MLEPLRLSQLTVALDARLIGEDAVFSAVSTDSRAITGPGOLFIAL
SGPRPDHDIYLAFAAKGAVAAVVEREVADAPLPOLLVDRTRAAALGRLGLNRKRTKG
PIAAVTGSSGKTAVKEMLASILRTQAGDAESVLATRGNNLNDGLVPLTLLQALPOHRS
ATVAVLNDKAEFTWKARASGRPLLTFSIDRPOADPRAADLQDARDAGCMGFRLOGVAG
EAOVLYNLGHRHNVAAAALAAAHALGPLDGIAGLQALQAPVKRAVAOQLTASGLR
VIDDSYNANPASMIAAIDILSGFSGRTVLVLDGMDGLSGMAEQHREVGAAYAAKYSVA
LVAVGLMAHAVAQAFGATGRHFADQASLIGALATEQPTTILIKSGRSAAMDKVVAAL
CGSSEESH"
2833..3915
/gene="murY"
2833..3915
/gene="murY"
/note="phospho-N-acetylmuramoyl-pentapeptide translocase"
/codon_start=1
/transl_table=11
/product="MurY"
/protein_id="AAG22121.1"
/db_xref="GI:10719754"
/translation="MLLLAEYLQOFYKGFYQYLTLRILSVLTALSLSLWLGPMW
IRTLQIPQIGQAVRNDGPOSHLSKQCTPTMGGALITATAISLTLLADLSNRVYVVL
VYTLLEFGAIGWDDYRKVIEKNSRGLPSRWKYPQWSYFGIAGAAYFLMYAETPIETL
IVPMKLSVEIQIGIFFVVLTYFVIGSSNAVNLTDGLDGLAIMPYVWAGAGIFCYL
SGNVKFAEYLLIPNPVAGELIVFCAALYAGLGLFWNTYPAQVPMGDVGALAGAA
LGTIAIVRQEIYLVFMGTVFVYMETLSVMIOVASFKLTGRRVFRMAPIHHEFLKQWP
DPRVIVRFWITVILVIGLATLKLK"
3929..5275
/gene="murD"
3929..5275
/gene="murD"
/note="UDP-N-acetylmuramyl-L-alanine:D-glutamate ligase;
recombinant enzyme catalyzed the ATP-dependent
addition of D-glutamate to the precursor sugar peptide"
/codon_start=1
/transl_table=11
/product="MurD"
/protein_id="AAG45237.1"
/db_xref="GI:12007718"
/translation="MSLIAADHFRIYVVLGKSGMSLVRLARRGLPFAVVDTRNPPE
LATLRAQYQVEARCEGLDAEFLCSARELYVSPGLSLRTPALVQAAKGVRSIGDIDL
FAREAKAPIAIVTGSNAKSTVTLVCEMAYADKRVAVGNGLTGTPALDILLADDIELYV
LELSPQLETCRLNAEAVATLVNSDEHMDRYDGMADYHLAKHRIPLRGARQVYVNRAD
ALTRPLIADTVPCFSGNLKPDFKAFGLIEEDQGWLAFOFGKLLPFGKLVGVVNGAHY
SNALAALGHAVAGLVPFDAMLGAKAFSGLAHRCQWVRERQGVSYVDDSKATNVGAAL
AAIEGLGADIDGKLVLGGDGKDFHOLREPVARFCRAVLLGLDRDAGLIAQALGNA
VPLVRVATLDEAVRQAELAREGDAVLLSPACASLDMFNFEERGLFAKAVEELA"
BASE COUNT      817 a 1781 c 1759 g 923 t
ORIGIN
Query Match      86.1%; Score 1248; DB 1; Length 5280;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1398; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY      2 GTGCTGATCGGCTCGCCACCTTGAAGCTGCGTTGAGACGAGAGAGATGAGCCTGAT 61
Db      3880 GTGCTGATCGGCTCGCCACCTTGAAGCTGCGTTGAGACGAGAGATGAGCCTGAT 3939
QY      62 CGCCTCCGACCACTTCGGCATGTTGTCGGCCTCGCAAGAGCGGCATGCTCCTGCTGG 121
Db      3940 CGCCTCCGACCACTTCGGCATGTTGTCGGCCTCGCAAGAGCGGCATGCTCCTGCTGG 3999
QY      122 CTACCTGGCGCGCGCGCTTGCCTTTCGCCCTGCTGTCGATACCCGAGAGAACCCGCGGA 181
Db      4000 CTACCTGGCGCGCGCGCTTGCCTTTCGCCCTGCTGTCGATACCCGAGAGAACCCGCGGA 4059
QY      182 GCTGCCCACTTCGTCGTCGCCAGTATCCGCAAGTGGAACTGCTGCGGCGAACTCGACGC 241
Db      4060 GCTGGCTACCTTCGTCGTCGCCAGTATCCGCAAGTGGAACTGCTGCGGCGAACTCGACGC 4119
QY      242 CGAGTTCCTCTGCTCGCGCGCGCGAACTCTATGTCAGCCCCCGCTTGTCTGCTGGCACCCC 301
Db      4120 CGAGTTCCTCTGCTCGCGCGCGCGAACTCTATGTCAGCCCCCGCTTGTCTGCTGGCACCCC 4179

```



QY	302	TGCGTGTGTACAGCGCGCGGAAAGGGGTGCGCATCTCCGGTGACATCGATCTCTTCCG	1382	AGAGGAGTAGCGTGATGCTG	1402
Db	4180	TGCGTGTGTACAGCGCGCGGAAAGGGGTGCGCATCTCCGGTGACATCGATCTCTTCCG	5260	AGAGGAGTAGCGTGATGCTG	5280
QY	362	CGCGAGGCGAAGGCCCGATCGTCCGCATCACCGGTTCCACAGCGAAGACCGTGAC	AE016779	301995 bp	DNA
Db	4240	CGCGAGGCGAAGGCCCGATCGTCCGCATCACCGGTTCCACAGCGAAGACCGTGAC	Pseudomonas putida KT2440	section 6 of 21	linear BCT 12-DEC-2002
QY	422	CACCTGTGTGGCGAAATGGCGGTGCGCGGACAGCGTGTGCGCGGCGCAACCT	AE016779	AE016779	
Db	4300	CACCTGTGTGGCGAAATGGCGGTGCGCGGACAGCGTGTGCGCGGCGCAACCT	AE016779.1	GI:26557023	
QY	482	CGGACCCCGCGGCTCGACCTGCTGGCGGACGACATCGAGCTGTAGTGTGGAGCTGTC	Pseudomonas putida KT2440		
Db	4360	CGGACCCCGCGGCTCGACCTGCTGGCGGACGACATCGAGCTGTAGTGTGGAGCTGTC	Pseudomonas putida KT2440		
QY	542	GAGCTTCAGCTGGAAACCTGGCATCGCTCAACCGCGAGGTGGGACCGTCTGAACGT	Bacteria: Proteobacteria; Gammaproteobacteria; Pseudomonadales;		
Db	4420	GAGCTTCAGCTGGAAACCTGGCATCGCTCAACCGCGAGGTGGGACCGTCTGAACGT	Pseudomonadaceae: Pseudomonas.		
QY	602	CAGCGAAGACCATATGGATCGCTACGAGCGCATGCTGACTACCACTGGCCAAACCG	1 (bases 1 to 301995)		
Db	4480	CAGCGAAGACCATATGGATCGCTACGAGCGCATGCTGACTACCACTGGCCAAACCG	Nelson, K., Paulsen, I., Weinel, C., Dodson, R., Hilbert, H., Fouts, D., Gill, S., Pop, M., Martins Dos Santos, V., Holmes, M., Brinkac, L., Beanan, M., DeBoy, R., Daugherty, S., Kolonay, J., Madupu, R., Nelson, W., White, O., Peterson, J., Khouri, H., Hance, I., Lee, P., Holtzapple, E., Scanlan, D., Tran, K., Moazzaz, A., Utterback, T., Rizzo, M., Lee, K., Kosack, D., Moesti, D., Wedler, H., Lauber, J., Hohseil, J., Straetz, M., Helm, S., Kiewitz, C., Eisen, J., Timmis, K., Dueterhoff, A., Tummier, B. and Fraser, C.		
QY	662	GATCTTCGCGGTGCCGCCAGGTGCTGTGAATCGCGCGATGCCCTGACCCGCGCT	Complete genome sequence and comparative analysis of the metabolically versatile Pseudomonas putida KT2440		
Db	4540	GATCTTCGCGGTGCCGCCAGGTGCTGTGAATCGCGCGATGCCCTGACCCGCGCT	Environ. Microbiol. 4 (12), 799-808 (2002)		
QY	722	GATCGCCGATCCGCTGCGTGTGCTGCTGGCGCTGAACAGCGCGACTTCAAGCTTT	2 (bases 1 to 301995)		
Db	4600	GATCGCCGATCCGCTGCGTGTGCTGCTGGCGCTGAACAGCGCGACTTCAAGCTTT	Nelson, K., Paulsen, I., Weinel, C., Dodson, R., Hilbert, H., Fouts, D., Gill, S., Pop, M., Martins Dos Santos, V., Holmes, M., Brinkac, L., Beanan, M., DeBoy, R., Daugherty, S., Kolonay, J., Madupu, R., Nelson, W., White, O., Peterson, J., Khouri, H., Hance, I., Lee, P., Holtzapple, E., Scanlan, D., Tran, K., Moazzaz, A., Utterback, T., Rizzo, M., Lee, K., Kosack, D., Moesti, D., Wedler, H., Lauber, J., Hohseil, J., Straetz, M., Helm, S., Kiewitz, C., Eisen, J., Timmis, K., Dueterhoff, A., Tummier, B. and Fraser, C.		
QY	782	CGGCTGTATCGAGGAGCGGCGAGTGGCTGCGCTTCCAGTTCGACAACTGCTGCC	Direct Submission		
Db	4660	CGGCTGTATCGAGGAGCGGCGAGTGGCTGCGCTTCCAGTTCGACAACTGCTGCC	Submitted (05-NOV-2002) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA		
QY	842	GTTTGGGCACTGAAGATCCGTGGCGGCCACAACTATTCCAAAGCGCTGCGCGCTGCC	Location/Qualifiers		
Db	4720	GTTTGGGCACTGAAGATCCGTGGCGGCCACAACTATTCCAAAGCGCTGCGCGCTGCC	1. .301995		
QY	902	GCTGGGCCATCGGCTCGGCTGCCCTGCGACGCCATGCTCGCGCGCTGAAGCGGTTTC	/organism="Pseudomonas putida KT2440"		
Db	4780	GCTGGGCCATCGGCTCGGCTGCCCTGCGACGCCATGCTCGCGCGCTGAAGCGGTTTC	/mol_type="genomic DNA"		
QY	962	CGGCTGCTCATCGCTGCGTGGTACCGCGCGGCGAGCGGCGTGAGCTACTACGACGA	/strain="KT2440"		
Db	4840	CGGCTGCTCATCGCTGCGTGGTACCGCGCGGCGAGCGGCGTGAGCTACTACGACGA	/db_xref="taxon:160488"		
QY	1022	TTCCAAGGCCAACACGTGCGCGCCGCCCTGCGCGCGATCGAGGGGCTGGTGGCGACAT	89. .712		
Db	4900	TTCCAAGGCCAACACGTGCGCGCCGCCCTGCGCGCGATCGAGGGGCTGGTGGCGACAT	/gene="sspA"		
QY	1082	CGACGGCAAGCTGTGCTGCTCGCGGCGGAGACGGCGAGCGCGGATTTCCATGACCT	/locus_tag="PP1320"		
Db	4960	CGACGGCAAGCTGTGCTGCTCGCGGCGGAGACGGCGAGCGCGGATTTCCATGACCT	/note="similar to GP:9864180; identified by sequence similarity; putative"		
QY	1142	GGCGAGCGGCTCGCGGCTTCTGCGGCGGTGTGCTGCTGCGGTGAGCGCGGCT	/codon_start=1		
Db	5020	GGCGAGCGGCTCGCGGCTTCTGCGGCGGTGTGCTGCTGCGGTGAGCGCGGCT	/transl_table=11		
QY	1202	GATTGCCAGGCACTGGGCAACCGCGTACCGCTGGTGGCGCTGCGCAAGCGCTGGAGGAAG	/product="stringent starvation protein A"		
Db	5080	GATTGCCAGGCACTGGGCAACCGCGTACCGCTGGTGGCGCTGCGCAAGCGCTGGAGGAAG	/protein_id="AA066944.1"		
QY	1262	AGTCGGAGCGCGCGAGCTGGCGCGGAGAGCGGATGCGGTGCTGTCGCCGCGCTG	/db_xref="GI:24982777"		
Db	5140	AGTCGGAGCGCGCGAGCTGGCGCGGAGAGCGGATGCGGTGCTGTCGCCGCGCTG	/translation="MGATNRLACYSADPDHYSHRVLRLVLAERKGVSVQILDVDPALHLPRIKAEVYPSGVTPLVDRLALYESTVMEYLEERYPHPPLMPYIPVARGNSRLLMHRIQDWCALADTFLDPRSSAEARKALRESLTGSPLSFSEFACFMSDQSLVDCCLLPTLRLPVLGIELPQAKPLLDYMERQFAREFQASLSVSVERMKL"		
QY	1322	CGCGAGCTTGGACATGTTCAAGAACTTCGAAAGACCGGACCGCTTTCGCCAAAGCCGT	693. .1154		
Db	5200	CGCGAGCTTGGACATGTTCAAGAACTTCGAAAGACCGGACCGCTTTCGCCAAAGCCGT	/gene="sspB"		
			/locus_tag="PP1321"		
			/note="similar to GB:S50157, GB:X60221, SP:P24539, and PID:509291; identified by sequence similarity; putative"		
			/codon_start=1		
			/transl_table=11		

```

/product="stringent starvation protein B"
/protein_id="AA066945.1"
/db_xref="GI:24982778"
/translations="MKASPKELMSSRPYLRYLYEWIVNDCTPHMLVNAEFPKV
QVDFGASQGVINLSPSAVSLHMDNDSEFGRFSGVAHSLFVPGAILGIYARE
NGQMFELEPFSMDGDEDDDDVQDDGPPDDGGPPRPTGRPSLKVVVK"
complement(1216..1794)
/locus_tag="pp1322"
complement(1216..1794)
/locus_tag="pp1322"
/codon_start=1
/transl_table=11
/product="lipoprotein, putative"
/protein_id="AA066946.1"
/db_xref="GI:24982779"
/translations="MPMRGLMALTLCLSVTCSSVLSTRNSPIEDDGRGTIGSK
IDSLETKAANVSKANPNLKGSHVYSSNGIVGLAQTPRADLKSIAEGTASQV
QRVKKVNEIQWQPSILARNDAWLTTKIKTOMLTDSAVPSRIKIVITENGIVYLL
GLVTRQEAATAVQGVSGVKIVLFEYID"
complement(1791..2378)
/gene="gmbA"
/locus_tag="pp1323"
complement(1791..2378)
/gene="gmbA"
/locus_tag="pp1323"
/notes="identified by match to PFAM protein family HMM
PF01380"
/codon_start=1
/transl_table=11
/product="phosphoheptose isomerase"
/protein_id="AA066947.1"
/db_xref="GI:24982780"
/translations="MQSIRLRFESAIETKQAMDILAPIHEQASLVNALLNEGKM
LACNGSAGDAQHSSELNRRERPSLPATLTDSTLFSIANDYSNVEFSKO
IRALGPGDVLIAISTSGNSANVIOQAHRDREMLVVALTGRDGGGMSALLPEDVE
IRVPSVTARIQVHLLAHCLCLDIDSLFGSEE"
complement(2471..2845)
/locus_tag="pp1324"
complement(2471..2845)
/locus_tag="pp1324"
/notes="identified by match to TIGR protein family HMM
TIGR00252"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein TIGR00252"
/protein_id="AA066948.1"
/db_xref="GI:24982782"
/translations="MMAASPTRAGAAETQALEYLOGGLOLLAARNWCKGGELDLV
MLDADTVFVEYRYRLHAGFGGALDSIDGRKQRLVLAATLFLQKEAHWGNYPCEFDV
VALQGGHHAGRPLOWLNKNAFEC"
complement(2842..4659)
/locus_tag="pp1325"
complement(2842..4659)
/locus_tag="pp1325"
/codon_start=1
/transl_table=11
/product="lipoprotein, putative"
/protein_id="AA066949.1"
/db_xref="GI:24982783"
/translations="MIACLRLLTALCLIAALLAACASSPSSSGLGELPRTDPAIEQLLE
KAASSAEDALLRLISAADLAYKQDFPRAARILQVPLDTLTKPAQVFASTLAHEL
AMRNQFKAALTAHPISQRLAEPLADQARTYSVHAALAEADGQALAAQOQRIILA
PLLGQAASNDNDIAWVASLPAEQLOQPTNDOTLAWGTSLAFVKSAGTLBOQAAA
IEAWVQKHDPHQAQPLALTKELASQPLTKIALLLPQEGPLAGVARNLRDGPMA
AHQAQSGGPAVQAVQFSDSRIGSLDDFYRQAAGVOLVIGPLEKPLVKOLAAPKO
LPITTLALYADAGKRAPPOLFOFGLAAEDEAREVARADGMVRAVALVPSGWDG
RVLAARQDWEQGGILLAAERIAQPVLAQQIADLEQLRQSEGRKSLQSTVGGSGIA
AQSRKQDIDFIFLASTPQAQIKPTLNFQYAGDVPVYATSNLISASGDIYNDMN
GIRFCFTWLLDTSNLRQVQVQWQPAAGSLGRLYAMGVDAYSLAPRLQKALPDN
RVILGSLGSLINANQRVERQLPWAIEFGSGGVKRLPDTVR"
complement(4837..5721)
/locus_tag="pp1326"
complement(4837..5721)

```

```

/locus_tag="pp1326"
/notes="identified by match to PFAM protein family HMM
PF00590"
/codon_start=1
/transl_table=11
/product="tetrahyptrole methylase family protein"
/protein_id="AA066950.1"
/db_xref="GI:24982784"
/translations="MCAVTDVAGASKSTVGLYVATPIGNLDDMSARALKVLADVAL
IAEDTRHSIRLLOHFGIDPPLAACHEHNERDEGGPRLSKLLAGENVALVSDAGTPLL
SDPGYHLVROARAGSVVPPGACALIAAASGALPSDRFIEGFLPAQAGRRARL
EQVREPRTLIFYEAPHRIDECLMEVFGDRPALARELTKFTETKGLPLAEUR
AFVAGDSNQQRGCVVLVVGWSAPEGQAISAEARQVLDLLLAELPLKRAALAAEIT
GVKNLLYQLALEKQKAE"
complement(5744..6164)
/locus_tag="pp1327"
/notes="This region contains a gene with one or more
premature stops or frame shifts, and is not the result of a
sequencing artifact; conserved hypothetical protein,
degenerate; similar to GB:X16944, and SP:P15488;
identified by sequence similarity; putative"
/locus_tag="pp1328"
complement(6455..6928)
/locus_tag="pp1328"
/notes="identified by match to PFAM protein family HMM
PF02381"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein TIGR00242"
/protein_id="AA066951.1"
/db_xref="GI:24982786"
/translations="MMGITAVFRGANVSLDAKRLAMPSTRYDELDSCNGOLIVTI
DAVDCLVVPLDEWEQIEAKRLPSLRENRLORLIGNAVLDELDCSGRFLVPP
RLPEYAKLDKAMLVQLNKFQLDWEDANVANSADLAAIQPCAMPDELRLIL"
complement(6925..7872)
/locus_tag="pp1329"
complement(6925..7872)
/locus_tag="pp1329"
/notes="similar to GP:9950652; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein TIGR00006"
/protein_id="AA066952.1"
/db_xref="GI:24982787"
/translations="MTIDSGFNHITVLLDEAVEALRADGCYLDGTFGRGHSRLIL
SKLPGOGRLLIGFDKQAIATGQALAAEDGRFVIVORSEALGAELVAAAGLKGKVCV
LLDLGVSSPOLDDPERGFSFLNDGPDLMRNPDQGVSAEFITATPVEETARVFEKG
EERFAGMARAVVREKOPKPTADLAELVKVANPAWEGKKNPATRFAOGLRIHVNN
ELGDLEAGLEAALDALEVGGRLAVISFHSLEDRIIVKLFMRKLVKGEADNLPRLPVQH
KVPEPKIKLIGKAQFAEALKANRSRSRAVRAEKLK"
complement(7869..8162)
/locus_tag="pp1330"
complement(7869..8162)
/locus_tag="pp1330"
/notes="similar to GB:M11186, GB:M62611, SP:P01178,
PID:189411, PID:386991, and PID:825699; identified by
sequence similarity; putative"

```

Query Match 3.0%; Score 44; DB 1; Length 301995;  
 Best Local Similarity 100.0%; Pred. No. 6.1e-10;  
 Matches 44; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 75 TTCCGCATCTTTCGCCCTCGGCAAGAGCGGATGTCCTCGT 118  
 |||||  
 Db 13867 TTCCGCATCTTTCGCCCTCGGCAAGAGCGGATGTCCTCGT 13910

RESULT 4  
 AX189058

```

LOCUS       AX189058                      1317 bp    DNA             linear       PAT 08-AUG-2001
DEFINITION   Sequence 259 from Patent WO0148209.
ACCESSION    AX189058
VERSION      AX189058.1  GI:15142764
KEYWORDS     Escherichia coli
SOURCE       Escherichia coli
ORGANISM     Escherichia coli
              Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
              Enterobacteriaceae; Escherichia.
REFERENCE    1
AUTHORS      Forsyth,R.A., Ohlsen,K.L. and Zyskind,J.W.
TITLES       Genes identified as required for proliferation of E. coli
JOURNAL      Patent: WO 0148209-A 259 05-JUL-2001;
              Elitra Pharmaceuticals, Inc. (US)
FEATURES     Location/Qualifiers
              1..1317
                /organism="Escherichia coli"
                /mol_type="genomic DNA"
                /db_xref="taxon:562"
              1..1317
                /note="unnamed protein product"
                /codon_start=1
                /transl_table=11
                /protein_id="CAC50739.1"
                /db_xref="GI:15142765"
                /translation="MADYQGNVVIIGLGLTGLSCVDFFLARGVTPRVMDTRMTPPGL
                DKLPAVERHTGSLNDEWMAADLIVASPGIALAHPSLSAADAGIEIVGDIELFCRE
                AQAPIVATGSGKSTVTLVGEAKAGVNVGGNIGLPAALMLDDECELYVLELS
                SFQLETTSSLAQVATILNVTEDHMDRYPFGLQQYRAAKLRIYENAKVCVYNADALT
                MPIRGADERCVSFGVMGDYHLNHQGETWLRVKEKLVNKKLSQHNVTALAA
                LALADAAGLPRASSLKALTTFTGLPHREYVLEHNGVRWINDSKATNVGSTEALNGL
                HVDGTLHLGLGDDGKSAFSPRLARYLNGDNVRLYCFGRDGAQLAALRPEVAEQETME
                QAMRLAPRVQPGDMVLLSPACASLDQKFNFEQRGNFARLAKELG"
              294 a 323 c 400 g 300 t
BASE COUNT   294 a 323 c 400 g 300 t
ORIGIN
Query Match      1.7%; Score 24; DB 6; Length 1317;
Best Local Similarity 100.0%; Pred. No. 5.2;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY  537 CTGTCGAGCTTCACCTGGAAC 560
    |||||||||||||||||||
Db   475 CTGTCGAGCTTCACCTGGAAC 498

RESULT 5
ECMURD      1629 bp  DNA             linear       BCT 12-SEP-1993
LOCUS       Escherichia coli murD gene for
DEFINITION   UDP-N-acetylmuramoyl-L-alanine:D-glutamate ligase (EC 6.3.2.9).
ACCESSION    X17609
VERSION      X17609.1  GI:42045
KEYWORDS     murD gene; peptidoglycan synthesis;
              UDP-N-acetylmuramoyl-L-alanine:D-glutamate ligase.
SOURCE       Escherichia coli
ORGANISM     Escherichia coli
              Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
              Enterobacteriaceae; Escherichia.
REFERENCE    1 (bases 1 to 1629)
AUTHORS      Mengin-Lecreulx,D. and van Heijenoort,J.
TITLES       Nucleotide sequence of the murD gene encoding the
JOURNAL      UDP-MurNAc-L-Ala-D-Glu synthetase of Escherichia coli
MEDLINE      Nucleic Acids Res. 18 (1), 183 (1990)
PUBMED       90174916
              2129548
              2 (bases 1 to 1629)
AUTHORS      Mengin-Lecreulx,D.
TITLES       Direct Submission
JOURNAL      Submitted (23-NOV-1989) Mengin-Lecreulx D., URA 1131 du CNRS,
              Biochimie Moleculaire et Cellulaire, Batiment 432, Universite Paris
              Sud, 91405 Orsay, France
COMMENT      *map: map position=2 minutes;
              Data kindly reviewed (26-FEB-1990) by Mengin-Lecreulx.

```

```

FEATURES     Location/Qualifiers
              1..1629
                /organism="Escherichia coli"
                /mol_type="genomic DNA"
                /strain="K12"
                /db_xref="taxon:562"
                212..216
                /note="ribosome binding site"
                227..1543
                /note="UDP-N-acetylmuramoyl-L-alanine:D-gluta ligase (AA
                1-438)"
                /codon_start=1
                /transl_table=11
                /protein_id="CAA35611.1"
                /db_xref="GI:42046"
                /db_xref="SWISS-PROT:P14900"
                /translation="MADYQGNVVIIGLGLTGLSCVDFFLARGVTPRVMDTRMTPPGL
                DKLPAVERHTGSLNDEWMAADLIVASPGIALAHPSLSAADAGIEIVGDIELFCRE
                AQAPIVATGSGKSTVTLVGEAKAGVNVGGNIGLPAALMLDDECELYVLELS
                SFQLETTSSLAQVATILNVTEDHMDRYPFGLQQYRAAKLRIYENAKVCVYNADALT
                MPIRGADERCVSFGVMGDYHLNHQGETWLRVKEKLVNKKLSQHNVTALAR
                VALADAAGLPRASSLKALTTFTGLPHREYVLEHNGVRWINDSKATNVGSTEALNGL
                HVDGTLHLGLGDDGKSAFSPRLARYLNGDNVRLYCFGRDGAQLAALRPEVAEQETME
                QAMRLAPRVQPGDMVLLSPACASLDQKFNFEQRGNFARLAKELG"
              351 a 403 c 486 g 389 t
BASE COUNT   351 a 403 c 486 g 389 t
ORIGIN
Query Match      1.7%; Score 24; DB 1; Length 1629;
Best Local Similarity 100.0%; Pred. No. 5;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY  537 CTGTCGAGCTTCACCTGGAAC 560
    |||||||||||||||||||
Db   701 CTGTCGAGCTTCACCTGGAAC 724

RESULT 6
ECMURD      2608 bp  DNA             linear       BCT 12-SEP-1993
LOCUS       Escherichia coli murD gene and ORF-Y (EC 6.3.2.9).
DEFINITION   X51584 X52117
ACCESSION    X51584.1  GI:42058
VERSION      X51584.1  GI:42058
KEYWORDS     membrane protein; murD gene;
              UDP-N-acetylmuramoyl-L-alanyl-D-glutamate synthetase; unidentified
              reading frame.
SOURCE       Escherichia coli
ORGANISM     Escherichia coli
              Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
              Enterobacteriaceae; Escherichia.
REFERENCE    1 (bases 1 to 2608)
AUTHORS      Ikeda,M., Wachi,M., Ishino,F. and Matsuhashi,M.
TITLES       Nucleotide sequence involving murD and an open reading frame ORF-Y
              spacing murf and ftsW in Escherichia coli
JOURNAL      Nucleic Acids Res. 18 (4), 1058 (1990)
MEDLINE      90192099
PUBMED       2179861
              2 (bases 1 to 2608)
AUTHORS      Ikeda,M.
TITLES       Direct Submission
JOURNAL      Submitted (02-JAN-1989) Ikeda M., Institute of Applied
              Microbiology, The University of Tokyo, 1-1-1 Yayoi, Bunkyo-ku,
              Tokyo 113, Japan
              3 (bases 26 to 1354)
AUTHORS      Flouret,B.
TITLES       Direct Submission
JOURNAL      Submitted (15-MAR-1990) Flouret B., URA 1131 du Centre National de
              la Recherche Scientifique, C N R S, Biochimie Moleculaire et
              Cellulaire, Batiment 432 Universite Paris-Sud, 91405 Orsay, France
COMMENT      Sequence independently determined by [3], but authors accept
              sequence [1] as definitive.
FEATURES     Location/Qualifiers
              1..2608
                /organism="Escherichia coli"

```



gene

EKLQHQHVDPSQENIVVQK"  
2107..3873

/gene="fts1"

/locus\_tag="SF0081"

2107..3873

/gene="fts1"

/locus\_tag="SF0081"

/note="Residues 1 to 588 of 588 are 99 pct identical to  
residues 1 to 588 of a 588 aa protein from Escherichia  
coli O157:H7 EDL933 ref: NP\_285780.1"

/codon\_start=1

/transl\_table=11

/product="septum formation; penicillin-binding protein 3;  
peptidoglycan synthetase"

/protein\_id="AA041745.1"

/db\_xref="GI:24050287"

/translation="MKAARKTKPKRQEHANFTISWRPALLCGCILLALAFELGRVAV  
LQVSPDLVREGMSRLRQVSTSRGMITDRSGPLAVSPVKAIWADPEKVEDAG  
GISVGRWALANALNIPDLQSLARINANPKRFTYLARQVNSOMADYIKRLKLPGRH  
LREESRRYPGSEVTAHLIGFTNVDSQIEGVKSFDFKWLGTQGPETVRKRDYRVI  
EDISSTDQAHNLSLSDERLQALVRELANAVAFKAESGSAVLVDVNTGEVLAMA  
NPSNPNLSGTSKPEAMRNTITDVEPGSTVAPMVVNTALQGVYRESVLNTIPY  
RINGHEIKDARISELTLGVQKNSVGVSKLALTPSSALVDYISFSGLGKATNUG  
LVGERSGLYPQKRWSDIERATFSFGYGLMVTPLQARVYATIGSYGTRPLSTKVD  
PPVPGERYFPESIVRTVVHMSVALPGGGKAAIKGYRIAIKRTAKKYVDPGRYI  
NKYIATAGVAPALPQRPALVVVINDPQAGKYVGAVSAPVFGAIMGVGLRTMNIETP  
ALTGTGKNEFVINDGEGTGGRS"

3860..5347

/gene="murE"

/locus\_tag="SF0082"

3860..5347

/gene="murE"

/locus\_tag="SF0082"

/EC\_number="6.3.2.13"

/note="Residues 1 to 495 of 495 are 99 pct identical to  
residues 1 to 495 of a 495 aa protein from Escherichia  
coli pir: S40595"

/codon\_start=1

/transl\_table=11

/product="UDP-N-acetylmuramoylalanyl-D-glutamate-2,

6-diaminopimelate ligase"

/protein\_id="AA041747.1"

/db\_xref="GI:24050288"

/translation="MADNRLRDLAPVPDPAPSRALREMTLDSRVAAAGDLFVAVVGH  
QADGRYIPQAIQGVAAIIEAKDEATDREMHGVPVYLSOLNERLSALAGRDFY  
HEPSDNLRLVGTGTNGKTTTOLLAQWSOLLGETSAVMGTGVNGLLKPIPTENTTG  
SNVDVQHELGLVGGATFCAMEVSSHGLVQHRVAALKFPAVSFTNLSRDHLDHGDM  
EYEAARKLLISEHCGQAIINADDEVGRWMLKLPDAVAVSMEDHIHPNCHGRWLKA  
TEVNYHDSGATIRFSSSSGDEIEHLMGFNFVSNLLALATLALGYPLADLLKTA  
RLQPCGMEVFTAFGKPTVVVYDHAHTDALEKALQAAHLCHAGKWLVCVCGGDRDK  
GKRPLMGAIIEEFADVAVVTDNPTPEPRALINDILAGMLDAGHAKVMEGRAEAVTC  
AVMQAKENDVVLVAGKHEDYQIVGNRLDYSDRVTVARLLGGIA"

5344..6702

/gene="murF"

/locus\_tag="SF0083"

5344..6702

/gene="murF"

/locus\_tag="SF0083"

/note="Residues 1 to 452 of 452 are 99 pct identical to  
residues 1 to 452 of a 452 aa protein from Escherichia  
coli K12 ref: NP\_414628.1"

/codon\_start=1

/transl\_table=11

/product="D-alanine:D-alanine-adding enzyme"

/protein\_id="AA041748.1"

/db\_xref="GI:24050289"

/translation="MISVTLSQLTDILNGLGQADITLDVTTDFRKLTPOGCLFVALK  
GERFDHAFDAQAGAGALVSRPLDIDLQILVTKDTRLAFGLAAWROQVPAVY  
VALTSGSKTSVKEMTAALISQCGNTLYTAGNLNDIGVPMTLRLLTPEYDYVTEIG  
ANHOCEIAWTVSLTPEALVNNLAAHLEGFGLAGVAKAKGEIFSGLPENGIAIMN  
ADNDRLNQSIVGSKVRWSPNANSDFATNLHVTSHGMEFTLOPTGSVDVLLP  
LPGRNINAAALAAALSMSVGTDLAIKAGLANLKAVPGRLPFIQIAENQLLLDDSYN  
ANVGSMTAAVQVLAEMPGRYRLVVVGDMALGAESEACHVQVGEAKAGIDRVLSMGK

gene

6696..7778

/gene="mraI"

/locus\_tag="SF0084"

6696..7778

/gene="mraI"

/locus\_tag="SF0084"

/note="Residues 1 to 360 of 360 are 100 pct identical to  
residues 1 to 360 of a 360 aa protein from Escherichia  
coli K12 ref: NP\_414629.1"

/codon\_start=1

/transl\_table=11

/product="phospho-N-acetylmuramoyl-pentapeptide

transferase?"

/protein\_id="AA041749.1"

/db\_xref="GI:24050290"

/translation="MLVWLAELVKYSGFNVSFLTRFRAIVSLTALFISLMMGPRM  
IAHLQKLSFGQVRNDGSPESHFSKRGTMTGGMILTAIVISLVLLWAYSPNYPVVCVL  
VVLGYGVIGVDDYRKVVKDDTKGLIARKYFWMSVIALGVAFALYLAGKDTPATQL  
VVPFKQVMPOLGLFYIILAVFVIVGTGNVNLDTGLDGLAIMPVTFVAGGFAIVAWA  
TGNMFASYLHPIPLRHAGELVIVCTAIVAGLGFELWENTYPAQVFMGDSGLAGGA  
LGITAVLQURQFELLVIMGVFVETLSVLVQLQVSPFKLRQGRIFRMAPIHHRHLYELGWP  
EPRTVTRFWIISLMLVLIGLATLKVR"

7781..9097

/gene="murD"

/locus\_tag="SF0085"

7781..9097

/gene="murD"

/locus\_tag="SF0085"

/note="Residues 1 to 438 of 438 are 99 pct identical to  
residues 1 to 438 of a 438 aa protein from Escherichia  
coli K12 ref: NP\_414630.1"

/codon\_start=1

/transl\_table=11

Query Match 1.7%; Score 24; DB 1; Length 11441;

Best Local Similarity 100.0%; Pred. No. 3.7;

Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 537 CTGTCGAGCTTCCAGCTGGAAACC 560

|||||

DB 8255 CTGTCGAGCTTCCAGCTGGAAACC 8278

RESULT 8

AE005185

LOCUS

DEFINITION

Escherichia coli O157:H7 EDL933 genome, contig 1 of 3, section 9 of

155.

ACCESSION AE005185 AE005174

VERSION AE005185.1 GI:12512782

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

Escherichia coli O157:H7 EDL933  
Escherichia coli O157:H7 EDL933  
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
Enterobacteriaceae; Escherichia.  
1 (bases 1 to 12791)  
Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glasner,J.D.,  
Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A.,  
Posfai,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L.,  
Grotbeck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamousis,K.,  
Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C.,  
Welch,R.A. and Blattner,F.R.

Genome sequence of enterohaemorrhagic Escherichia coli O157:H7

Nature 409 (6819), 529-533 (2001)

21074935

MEDLINE 11206551

PUBMED

REFERENCE 2 (bases 1 to 12791)

Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glasner,J.D.,

Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A.,

Posfai,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L.,

Grotbeck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamousis,K.,

Grotbeck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamousis,K.,

Apodaca, J., Anantharaman, T.S., Lin, J., Yen, G., Schwartz, D.C.,  
 Welch, R.A. and Blattner, F.R.  
 TITLE Direct Submission  
 JOURNAL Submitted (22-OCT-2000) Laboratory of Genetics, University of  
 Wisconsin, 445 Henry Mall, Madison, WI 53706, USA

FEATURES  
 source location/Qualifiers  
 1. .12791  
 /organism="Escherichia coli O157:H7 EDL933"  
 /mol\_type="genomic DNA"  
 /strain="EDL933"  
 /serotype="O157:H7"  
 /db\_xref="taxon:155864"  
 /note="enterohemorrhagic"  
 72. .1076  
 /gene="fruR"  
 /note="synonym: Z0090"  
 72. .1076  
 /gene="fruR"  
 /function="regulator; Energy metabolism, carbon:  
 Glycolysis"  
 /notes="Residues 1 to 334 of 334 are 100.00 pct identical  
 to residues 1 to 334 of 334 from Escherichia coli K-12  
 Strain MG1655: B0080"  
 /codon\_start=1  
 /transl\_table=11  
 /product="transcriptional repressor of fru operon and  
 others"  
 /protein\_id="AAG54384.1"  
 /db\_xref="GI:12512783"  
 /translation="MKLDEIARLGVSRRTASVINGKAKQVRVSKTVEKVMVAVRE  
 HNYHPNVAAGRAGRTSGILVIPLENTSYRIANYLERQARQGYQLIACSEDQ  
 PNEMRCIEHLQRQVDIAITVSLPEHPFYQRWANDPFPPIVALDRALDRHFTSV  
 GADODAEMLKREPAETLVYLGLALPELSVSLREOGFTAKWKDDPREVHFLYAN  
 SYDERAAQLFEKWLTHPMQALFTTSEALQGVMDVTLRRDGKLPDLATATFGDN  
 ELLDFQCPLVAVQARHDAERVAERVLIVLASLDPEKPKPGTTRIARNLYRKGVLRS  
 1678. .2136  
 /gene="yabB"  
 /note="synonym: Z0091"  
 1678. .2136  
 /gene="yabB"  
 /function="orf; Unknown function"  
 /note="Residues 1 to 152 of 152 are 99.34 pct identical to  
 residues 1 to 152 of 152 from Escherichia coli K-12 Strain  
 MG1655: B0081"  
 /codon\_start=1  
 /transl\_table=11  
 /product="orf, hypothetical protein"  
 /protein\_id="AAG54385.1"  
 /db\_xref="GI:12512784"  
 /translation="MFRGATLVNLSKGRSLVPTRYREOLLENAAGQMVCTIDIHPC  
 LLLYPLPEWEITEOKLSRLSSMNPVRRVORLLIGHASECONDGAGRLLIAPVLRQHA  
 GLTKEWLVGQPKFELWDETTHQVQKEDIDAEQLATGDLSERLQDLSL"  
 2138. .3079  
 /gene="yabC"  
 /note="synonym: Z0092"  
 2138. .3079  
 /gene="yabC"  
 /function="orf; Not classified"  
 /note="Residues 1 to 313 of 313 are 100.00 pct identical  
 to residues 1 to 313 of 313 from Escherichia coli K-12  
 Strain MG1655: B0082"  
 /codon\_start=1  
 /transl\_table=11  
 /product="putative apolipoprotein"  
 /protein\_id="AAG54386.1"  
 /db\_xref="GI:12512785"  
 /translation="WMENYKHTVTLLDEAVNGLINRPDGIYIDGTFGRGSHRLLSQ  
 LGEGRLLAIDRQPAIAVAKTIDDPFSIIHGPFSAIGEYVAERDLKIDGILIDL  
 GYSSQLDAAERGFSDMRDGLMDRPTRGQSAAEWLQTAEEADIAWLKTYGERF  
 AKRIARIVERNREOQPMRTKELAEVVAATPVKDKFKPATRTFOAVRIWNSLEE  
 IEQALKSSINLVAPGRSLIISFSLSDRIYKFRNRSRGQVPAGLPMTQEQLKL  
 GGRQLRALGKLPGEVEVAENPRAKSSVLRIARTNA"

gene  
 3076. .3441  
 /gene="ftsL"  
 /note="synonym: Z0093"  
 3076. .3441  
 /gene="ftsL"  
 /function="phenotype; Cell division"  
 /note="Residues 1 to 121 of 121 are 100.00 pct identical  
 to residues 1 to 121 of 121 from Escherichia coli K-12  
 Strain MG1655: B0083"  
 /codon\_start=1  
 /transl\_table=11  
 /product="cell division protein; ingrowth of wall at  
 septum"  
 /protein\_id="AAG54387.1"  
 /db\_xref="GI:12512786"  
 /translation="MISRYTEALSKVSGMSGSHERALPCVIGDDLLRECKLPCLCFI  
 CILITAVVYVTTAHTRLRLTAQREQLVLENDALDIEWRNLILEENALGDHSRVERIAT  
 EKLQMQHVDPSQENIVVQK"  
 3457. .5223  
 /gene="ftsI"  
 /note="synonym: Z0094"  
 3457. .5223  
 /gene="ftsI"  
 /function="enzyme; Cell division"  
 /note="Residues 1 to 588 of 588 are 100.00 pct identical  
 to residues 1 to 588 of 588 from Escherichia coli K-12  
 Strain MG1655: B0084"  
 /codon\_start=1  
 /transl\_table=11  
 /product="septum formation; penicillin-binding protein 3;  
 peptidoglycan synthetase"  
 /protein\_id="AAG54388.1"  
 /db\_xref="GI:12512787"  
 /translation="MKAATQKPKQKEEHANFISHRFALLCCILLALAFLLGRVAN  
 LQVYSDMLVKEGDMRSLRQVQSTSRGMITDRSGRLAVSPVPKAIADPKRVHONG  
 GTSVGRKALANALNIPDLQSARINANPKGRFYIARQVNPDMADYIKKKLPGIHL  
 LREESRYYPGEVTAHLIGFTNVDSQEGVEKSPDKWLTGQGERIVKRDYGRVI  
 EDISDSDQAHLNALSIDERLQALVYRELNNAAFKAESGSAVLVDVNTGVLAMA  
 NSPSYNNLSGTPKRAMRNTITDVFEPGSTVKPMVNTALQGVVRENSVLNTIPY  
 RINGHEIKDVARYSLETLTGVLQKSSNVGSKLAMPSSALVDVTVSRGLGKATNLG  
 LVGERSGLYPQKQWSDIERATFSFGYGLMVTPLQARVATIGSGIYRPLSITKVD  
 PVPGRVFPPEISIVTVHMMESVALPGGGGVKAAIKGYRIAKTAKRVGPDGGRYI  
 NIYIATAGVAPASQPRFALVVVINDPQAGKYGGAVSAPVFGAINGGVLRTMNIEPD  
 ALLTGDKNEFVINGQSGTGGRS"  
 5210. .6697  
 /gene="mure"  
 /note="synonym: Z0095"  
 5210. .6697  
 /gene="mure"  
 /function="enzyme; Cell envelop: Murein sacculus,  
 peptidoglycan"  
 /note="Residues 1 to 495 of 495 are 98.98 pct identical to  
 residues 1 to 495 of 495 from Escherichia coli K-12 Strain  
 MG1655: B0085"  
 /codon\_start=1  
 /transl\_table=11  
 /product="meso-diaminopimelate-adding enzyme"  
 /protein\_id="AAG54389.1"  
 /db\_xref="GI:12512788"  
 /translation="MADRNLRLDLPWPDPAPSRALREMTLDSRVAAGDLFVAVVGH  
 QADGRYIPQAIAGQVAAIAEAKGEATDGEIREMHGVPIYLSQLNERLSALAGRFY  
 HPSDNLXXVGTGTNGTKTTTQLLAQNSQLLGETSAVNTGNGLLGRVITPTNTTG  
 SAVDVOHELAVLQDQDQATFCAMEVSSHGLVQHRVAALKFAASVFTNLSRDHLDYHGD  
 EHYEAKWLLYSEHHCGOAIINADDEVRRLAKLPDAVAVSMEDHINPCNGLRWLKA  
 TEVNVHDSGATIRFSSSGDGEIESHLMGAEVNNLLLATLALLGYPLADLLKTA  
 RLQPCGRMEVETAPGKPTVVVDYATHPDALEKALQARLHCAGKLCWCVGCGGDRDK  
 GKRRLMGAEEEFADVAVVYTDNPTPEPRALINDILAGMLDAGHAKVMEDRAEAVTC  
 AYMQAKENDEVLVAGKHEDYQIVGNQRLDYSDRVTVARLLGVIA"  
 6694. .8052  
 /gene="murf"  
 /note="synonym: Z0096"  
 6694. .8052  
 /gene="murf"



protein\_bind /note="central position to predicted promoter: -319.5"  
 554. .565 /bound\_moiety="fyrR predicted site"  
 /note="central position to predicted promoter:64"  
 /bound\_moiety="Arac predicted site"  
 646. .1767 /gene="leuo"  
 /note="synonym: b0076"  
 646. .1767 /gene="leuo"  
 /note="synonym: b0076"  
 /function="putative regulator; Amino acid biosynthesis: Leucine"  
 /note="o373; 100 pct identical to LEUO\_ECOLI SW: P10151(290 aa) but contains 59 additional N-ter aa and 24 C-term residues; 100 pct identical to PIR: S40589 but contains 24 additional C-term residues"  
 /codon\_start=1  
 /transl\_table=11  
 /product="probable transcriptional activator for leuABCD operon"  
 /protein\_id="AAC73187.1"  
 /db\_xref="GI:1786264"  
 /translation="WTHSTAMDSVFTRIFMFSEYSEFCFFLFYMHDKSYSSGLFLC  
 IPIRELSVTVELSMPEVOTDHPETAELSKPOLRMVDLNLTLTFDVAHQEONITRAA  
 HVLGMSQPAVNAVARLKYMFNDELFRYGRGIQPTARAFQFLGVSQRQAQLQVNELP  
 GSGFEPASSERFHLVCSPFLDSILTSQIYNHIEQAPNIHVHFKSSLNQNTHEQLRY  
 QTEFEVISEDHRPEFTSVPLFKDEMVLVASKNHPTIKGPLLKHVDVNEQHAASLD  
 RPAFSQPAVDVVKQASTAYOGMAKMSVLSVVSOPHLVLAIPRWLAEEFAESLELQV  
 LPLPLKQNSRTCYLSHWEAAGRKQHWEEQLVLSICKR"  
 1802. .1816 /note="central position to predicted promoter: -39.5"  
 /bound\_moiety="Lrp predicted site"  
 1813. .1839 /note="factor Sigma70; promoter ilvIHp4; documented+1 at  
 85394"  
 1833. .1847 /note="central position to predicted promoter: -8.5"  
 /bound\_moiety="Lrp predicted site"  
 1838. .1866 /note="factor Sigma70; promoter ilvIHp3; documented+1 at  
 85420"  
 1915. .1929 /note="central position to predicted promoter:73.5"  
 /bound\_moiety="Lrp predicted site"  
 1949. .1963 /note="central position to predicted promoter:107.5"  
 /bound\_moiety="Lrp predicted site"  
 1955. .1984 /note="factor Sigma70; promoter ilvIHp2; documented+1 at  
 85534"  
 1978. .1992 /note="central position to predicted promoter:136.5"  
 /bound\_moiety="Lrp predicted site"  
 1995. .3809 /gene="ilvi"  
 /note="synonym: b0077"  
 1995. .3809 /gene="ilvi"  
 /EC\_number="4.1.3.18"  
 /function="enzyme; Amino acid biosynthesis: Isoleucine, Valine"  
 /note="o504; 98 pct identical (1 gap) to 522 residues from ILVI\_ECOLI SW: P00893 (566 aa)"  
 /codon\_start=1  
 /transl\_table=11  
 /product="acetolactate synthase III, valine sensitive, large subunit"  
 /protein\_id="AAC73188.1"  
 /db\_xref="GI:1786265"  
 /translation="MFMSLAANCLSKIGRLCMCTFFHFRQTVQAMEMLSGAEMVVRSL  
 IDQGVKQVFGYGGADLDYDALHTVGGIDHVLVRHEQAAVHMADCLARATGCEGVVL  
 VTSGGATNATIGIATAYNDSIPLVLSQVATSLIGYDAFOECDNVGISRVPVKHSF  
 LVKQTEDIQVLKKAFLWAASGRPGVPVVDLPDLILNPANKLPYVWPESVSMRSYNT

TTGKQGIKRALQTLVAAKPPVYVGGGATAGCQHQKLKTVLALNLVVCVSLMGLGA  
 FPAHQALGLMGHGTYEANTMHNADVIFAVGVRFDRTTNNLAKYCPNATVLRHID  
 IDPTSISKVTADIPVGDARQVLEOMLELLSOEASHQPLDEIRDMWQOIEQARQOC  
 LKYDTHSEKIKPOAVIETLWRLFKGDAYVTSVYGQHQMFALYYPDKPRMINSGL  
 GTMGFGPALGVMALPEETVVCVTGDSIQMNIQELSTALQYELPVLVYNNLRFL  
 GMVKQODMIYSGRHSQSYNQLPDPFVRLAEAYGHVGIQISHPHELESKLEALEQVR  
 NNRLVFVDVTVDGSEHVYPMQIRGGMDENWLSKTRT"  
 1998. .2012 /gene="ilvi"  
 /note="central position to predicted promoter:156.5"  
 /bound\_moiety="Lrp predicted site"  
 2017. .2045 /gene="ilvi"  
 /note="factor Sigma70; promoter ilvIHp1; documented+1 at  
 85597"  
 3812. .4303 /note="synonym: b0078"  
 /gene="ilvH"  
 3812. .4303 /note="synonym: b0078"  
 /EC\_number="4.1.3.18"  
 /EC\_number="4.1.3.18"  
 Query Match 1.7%; Score 24; DB 1; Length 21757;  
 Best Local Similarity 100.0%; Pred. No. 3.4;  
 Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps -0;  
 QY 537 CTGTCGAGCTTCCAGCTGGAAACC 560  
 |||||  
 Db 14016 CTGTCGAGCTTCCAGCTGGAAACC 14039  
 |||||  
 RESULT 10  
 AX191720  
 LOCUS AX191720 28277 bp DNA linear PAT 15-AUG-2001  
 DEFINITION Sequence 2 from Patent WO0149775.  
 ACCESSION AX191720  
 VERSION AX191720.1 GI:15209889  
 KEYWORDS  
 SOURCE Escherichia coli  
 ORGANISM Escherichia coli  
 Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 Enterobacteriaceae; Escherichia.  
 REFERENCE 1  
 AUTHORS Iversen,P.L.  
 TITLE Antisense antibacterial cell division composition and method  
 JOURNAL Patent: WO 0149775-A 2 12-JUL-2001;  
 AVI Biopharma, Inc. (US)  
 FEATURES  
 source Location/Qualifiers  
 1. .28277  
 /organism="Escherichia coli"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:562"  
 BASE COUNT 6714 a 6725 c 7993 g 6845 t  
 ORIGIN  
 Query Match 1.7%; Score 24; DB 6; Length 28277;  
 Best Local Similarity 100.0%; Pred. No. 3.2;  
 Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 537 CTGTCGAGCTTCCAGCTGGAAACC 560  
 |||||  
 Db 14091 CTGTCGAGCTTCCAGCTGGAAACC 14114  
 |||||  
 RESULT 11  
 EC2MIN  
 ID EC2MIN standard; DNA; PRO: 28277 BP.  
 XX  
 XX X55034; M10429;  
 XX AC  
 XX SV X55034.1  
 XX SV  
 Dt 21-FEB-1991 (Rel. 27, Created)



DT 05-JUL-1999 (Rel. 60, Last updated, Version 31)  
XX E. coli 2 minute region  
KW ddl gene; envA gene; ftsA gene; ftsQ gene; ftsW gene; ftsZ gene; ilvH gene;  
KW ilvI gene; leuA gene; leuO gene; mraY gene; murC gene; murD gene;  
KW mre gene; murF gene; murG gene; mutT gene; orfA; orfB; orfC; orfX;  
KW pbbB gene; secA gene; shi gene.  
XX  
OS Escherichia coli  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Escherichia.  
XX  
RN [1]  
RP 1-28277  
RA Ayala J.A.;  
RT  
RL Submitted (08-JAN-1991) to the EMBL/GenBank/DBJ databases.  
RL Ayala J.A., Instituto de Biologia Molecular, Centro de Biologia Molecular,  
RL Universidad Autonoma, Canto-Bianco 28049, Madrid, Spain.  
XX  
RN [2]  
RA Ayala J.A.;  
RT "Regulation of transcription at the 2-minute region of the genetic map of  
RT Escherichia coli";  
RT Unpublished.  
XX  
RN [3]  
RP 1-306  
RX MEDLINE; 82078077.  
RX PUBMED; 6171647.  
RA Wessler S.R., Calvo J.M.;  
RT "Control of leu operon expression in Escherichia coli by a transcription  
RT attenuation mechanism";  
RL J. Mol. Biol. 149(4):579-597(1981).  
XX  
RN [4]  
RP 268-1130  
RX MEDLINE; 86223773.  
RX PUBMED; 3519576.  
RA Haughn G.W., Wessler S.R., Gemmill R.M., Calvo J.M.;  
RT "High A + T content conserved in DNA sequences upstream of leuABCD in  
RT Escherichia coli and Salmonella typhimurium";  
RL J. Bacteriol. 166(3):1113-1117(1986).  
XX  
RN [5]  
RP 843-1812  
RX MEDLINE; 88320486.  
RX PUBMED; 3413113.  
RA Henikoff S., Haughn G.W., Calvo J.M., Wallace J.C.;  
RT "A large family of bacterial activator proteins";  
RL Proc. Natl. Acad. Sci. U.S.A. 85(18):6602-6606(1988).  
XX  
RN [6]  
RP 1799-2187  
RX MEDLINE; 85234358.  
RX PUBMED; 3891724.  
RA Haughn G.W., Squires C.H., Defelice M., Largo C.T., Calvo J.M.;  
RT "Unusual organization of the ilvIH promoter of Escherichia coli";  
RL J. Bacteriol. 163(1):186-198(1985).  
XX  
RN [7]  
RP 2101-4431  
RX MEDLINE; 83272971.  
RX PUBMED; 6308579.  
RA Squires C.H., Defelice M., Devereux J., Calvo J.M.;  
RT "Molecular structure of ilvIH and its evolutionary relationship to ilvG in  
RT Escherichia coli";  
RL Nucleic Acids Res. 11(15):5299-5313(1983).  
XX  
RN [8]  
RP 4274-6093  
RX MEDLINE; 90330585.  
XX  
RX PUBMED; 2198273.  
RA Leclerc G., Noel G., Drapeau G.;  
RT "Molecular cloning, nucleotide sequence and expression of shi, a new gene  
RT in the 2-minute region of the genetic map of Escherichia coli";  
RL J. Bacteriol. 172(8):4696-4700(1990).  
XX  
RN [9]  
RP 6088-7587  
RX MEDLINE; 90251464.  
RA Gomez M.J., Fluoret B., Van Heijenoort J., Ayala J.A.;  
RT "Nucleotide sequence of the regulatory region of pbbB gene of Escherichia  
RT coli";  
RL Nucleic Acids Res. 18:2813-2813(1990).  
XX  
RN [10]  
RP 7316-10074  
RX MEDLINE; 83296957.  
RX PUBMED; 6350821.  
RA Nakamura M., Maruyama I.N., Soma M., Kato J.I., Suzuki H., Hirota Y.;  
RT "On the process of cellular division in Escherichia coli: nucleotide  
RT sequence of the gene for penicillin-binding protein 3";  
RL Mol. Gen. Genet. 191(1):1-9(1983).  
XX  
RN [11]  
RX MEDLINE; 90124047.  
RX PUBMED; 2692800.  
RA Tao J.S., Ishiguro E.E.;  
RT "Nucleotide sequence of the murE gene of Escherichia coli";  
RL Can. J. Microbiol. 35(11):1051-1054(1989).  
XX  
RN [12]  
RX MEDLINE; 90328986.  
RX PUBMED; 2198024.  
RA Michaud C., Parquet C., Flouret B., Blanot D., Van Heijenoort J.;  
RT "Revised interpretation of the sequence containing the mreE gene encoding  
RT the UDP-N-acetylmuramyl-tripeptide synthetase of Escherichia coli";  
RL Biochem. J. 269(1):277-278(1990).  
XX  
RN [13]  
RP 11142-12634  
RX MEDLINE; 89345095.  
RA Parquet C., Fluoret B., Menguin-Lecreulx D., Van Heijenoort J.;  
RT "Nucleotide sequence of the murF gene encoding the UDP-MurNac- pentapeptide  
RT synthetase of Escherichia coli";  
RL Nucleic Acids Res. 17:5379-5379(1989).  
XX  
RN [14]  
RP 12423-15030  
RX MEDLINE; 90192099.  
RA Ikeda M., Wachi M., Ishino F., Matsuhashi M.;  
RT "Nucleotide sequence involving murD and an open reading frame orf-Y spacing  
RT murF and ftsW in Escherichia coli";  
RL Nucleic Acids Res. 18:1058-1058(1990).  
XX  
RN [15]  
RP 13392-15020  
RX MEDLINE; 90036736.  
RA Menguin-Lecreulx D., Van Heijenoort J.;  
RT "Nucleotide sequence of the murD gene encoding the UDP-MurNac- L-Ala-D-gly  
RT synthetase of Escherichia coli";  
RL Nucleic Acids Res. 18:183-183(1990).  
XX  
RN [16]  
RP 14743-16239  
RX MEDLINE; 90036736.  
RX PUBMED; 2509435.  
RA Ikeda M., Sato T., Wachi M., Jung H.K., Ishino F., Kobayashi Y.,  
RA Matsuhashi M.;  
RT "Structural similarity among Escherichia coli FtsW and RodA proteins and  
RT Bacillus subtilis SpoVE protein, which function in cell division, cell  
RT elongation, and spore formation, respectively";  
RL J. Bacteriol. 171(11):6375-6378(1989).  
XX  
RN [17]

RP 16094-17806  
RA Menguin-Lecreux D., Textier L., Van Heijenoort J.;  
RT "Nucleotide sequence of the cell-envelope murG gene of *Escherichia coli*";  
RL Nucleic Acids Res. 18:2810-2810(1990).  
XX [18]  
RN 16094-18886  
RP MEDLINE; 90326550.  
RX Ikeda M., Wachi M., Jung H.K., Ishino F., Matsushashi M.;  
RA "Nucleotide sequence of the cell-envelope murG and murC in the *mra* gene cluster region  
RT of *Escherichia coli*";  
RL Nucleic Acids Res. 18:4014-4014(1990).  
XX [19]  
RN 19464-21952  
RP MEDLINE; 85054557.  
RX PUBMED; 6094474.  
RA Robinson A.C., Kenan D.J., Hatfull G.F., Sullivan N.F., Spiegelberg R.,  
RA Donachie W.D.;  
RT "DNA sequence and transcriptional organization of essential cell division  
RT genes *ftsQ* and *ftsA* of *Escherichia coli*: evidence for overlapping  
RT transcriptional units";  
RL J. Bacteriol. 160(2):546-555(1984).  
XX [20]  
RN 21464-23333  
RP MEDLINE; 86083166.  
RX PUBMED; 3000876.  
RA Yi Q.M., Lutkenhaus J.;  
RT "The nucleotide sequence of the essential cell-division gene *ftsZ* of  
RT *Escherichia coli*";  
RL Gene 36(3):241-247(1985).  
XX [21]  
RN 22964-25011  
RP MEDLINE; 88058745.  
RX PUBMED; 2824434.  
RA Beall B., Lutkenhaus J.;  
RT "Sequence analysis, transcriptional organization, and insertional  
RT mutagenesis of the *envA* gene of *Escherichia coli*";  
RL J. Bacteriol. 169(12):5408-5415(1987).  
XX [22]  
RN 23989-27799  
RP MEDLINE; 88298644.  
RX PUBMED; 2841285.  
RA Schmidt M., Rollo E., Grodberg I., Oliver D.;  
Query Match 1.7%; Score 24; DB 15; Length 28277;  
Best Local Similarity 100.0%; Pred. No. 3.2;  
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 537 CTGTCGAGCTTCAGCTGGAAC 560  
|||||  
Db 14091 CTGTCGAGCTTCAGCTGGAAC 14114  
RESULT 12  
LOCUS ECOL110K  
DEFINITION *Escherichia coli* genomic DNA.(0.0 - 2.4 min).  
ACCESSION D10483 J01597 J01863 J01706 K01298 K01990 M10420 M10611 M12544  
VERSION V00259 X04711 X54847 X54945 X55034 X56742  
KEYWORDS Complete and shotgun sequencing; *thrA*; *thrAl*; *thrA2*; *thrB*; *thrC*;  
*yaaA*; *yaaJ*; *talB*; *mog*; *chlG*; *dnaK*; *dnaJ*; *groP*; *gef*; *nhaA*; *ant*;  
*tpst*; *yaaC*; *ileS*; *lspA*; *yaaB*; *lytB*; *yaaF*; *dapB*; *carA*; *pyrA*; *carB*;  
*yaaV*; *calE*; *calD*; *calT*; *calA*; *calT*; *fixA*; *fixC*; *yaaT*; *yabE*;  
*yabF*; *kefC*; *trkC*; *trkQ*; *folA*; *tmrA*; *apaH*; *apaG*; *pdxA*; *suaR*; *imp*; *ostA*;  
*yabH*; *yabK*; *yabQ*; *hepA*; *polB*; *dina*; *araD*; *araB*; *araC*; *yabI*;  
*yabJ*; *yabK*; *tbpA*; *yabN*; *leuD*; *leuC*; *leuB*; *leuA*; *leuP*; *lucO*;  
*livI*; *livH*; *brnP*; *shl*; *frur*; *yabB*; *yabC*; *ftsL*; *ftsI*; *murE*;  
*murF*; *mra*; *mraY*; *murX*; *murD*; *ftsW*; *murC*; *ddl*; *ddlB*; *ftsQ*;

*ftsA*; *divA*; *ftsZ*; *sfirA*; *sulB*; *lpxC*; *envA*; *asmB*; *yacA*; *seca*; *prlD*;  
*azi*; *pea*; *mutT*; *yacG*; *htgA*; *htp*; *yaaI*; *nhaR*; *antO*; *calF*; *fixB*;  
*yaaU*; *yabO*.  
*Escherichia coli*  
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
Enterobacteriaceae; *Escherichia*.  
1  
Ohtsubo, H. and Ohtsubo, E.  
Nucleotide sequence of an insertion element, ISI  
Proc. Natl. Acad. Sci. U.S.A. 75 (2), 615-619 (1978)  
78137003  
PUBMED 273224  
2  
Smith, B.R. and Schleif, R.  
Nucleotide sequence of the L-arabinose regulatory region of  
*Escherichia coli* K12  
J. Biol. Chem. 253 (19), 6931-6933 (1978)  
79005683  
PUBMED 357433  
3  
Greenfield, L., Boone, T. and Wilcox, G.  
DNA sequence of the *araBAD* promoter in *Escherichia coli* B/r  
Proc. Natl. Acad. Sci. U.S.A. 75 (10), 4724-4728 (1978)  
79116194  
PUBMED 368797  
4  
Johnsrud, L.  
DNA sequence of the transposable element ISI  
Mol. Gen. Genet. 169 (2), 213-218 (1979)  
79177885  
PUBMED 375010  
5  
Smith, D.R. and Calvo, J.M.  
Nucleotide sequence of the *E coli* gene coding for dihydrofolate  
reductase  
Nucleic Acids Res. 8 (10), 2255-2274 (1980)  
81053692  
PUBMED 6159575  
6  
Ogden, S., Haggerty, D., Stoner, C.M., Kolodrubetz, D. and Schleif, R.  
The *Escherichia coli* L-arabinose operon: binding sites of the  
regulatory proteins and a mechanism of positive and negative  
regulation  
Proc. Natl. Acad. Sci. U.S.A. 77 (6), 3346-3350 (1980)  
81013881  
PUBMED 6251457  
7  
Katinka, M., Cossart, P., Sibilli, L., Saint-Girons, I.,  
Chaivignac, M.A., Le Bras, G., Cohen, G.N. and Yaniv, M.  
Nucleotide sequence of the *thrA* gene of *Escherichia coli*  
Proc. Natl. Acad. Sci. U.S.A. 77 (10), 5730-5733 (1980)  
81077247  
PUBMED 7003595  
8  
Miyada, C.G., Horwitz, A.H., Cass, L.G., Tinko, J. and Wilcox, G.  
DNA sequence of the *araC* regulatory gene from *Escherichia coli* B/r  
Nucleic Acids Res. 8 (22), 5267-5274 (1980)  
81124262  
PUBMED 7008027  
9  
Cossart, P., Katinka, M. and Yaniv, M.  
Nucleotide sequence of the *thrB* gene of *E. coli*, and its two  
adjacent regions; the *thrAB* and *thrBC* junctions  
Nucleic Acids Res. 9 (2), 339-347 (1981)  
81150470  
PUBMED 6259626  
10  
Lee, N.L., Gielow, W.O. and Wallace, R.G.  
Mechanism of *araC* autoregulation and the domains of two overlapping  
promoters, *PC* and *PBAD*, in the L-arabinose regulatory region of  
*Escherichia coli*  
Proc. Natl. Acad. Sci. U.S.A. 78 (2), 752-756 (1981)  
JOURNAL

MEDLINE  
 PUBMED  
 REFERENCE  
 11  
 AUTHORS  
 TITLE  
 Nucleotide sequence of the gene for ribosomal protein S20 and its  
 flanking regions  
 J. Biol. Chem. 256 (15), 8177-8182 (1981)  
 MEDLINE  
 PUBMED  
 REFERENCE  
 12  
 AUTHORS  
 TITLE  
 Is the amino acid but not the nucleotide sequence of the  
 Escherichia coli *araC* gene conserved?  
 J. Mol. Biol. 154 (4), 649-652 (1982)  
 MEDLINE  
 PUBMED  
 REFERENCE  
 13  
 AUTHORS  
 TITLE  
 Sequence of the *malK* gene in *E. coli* K12  
 Nucleic Acids Res. 10 (22), 7449-7458 (1982)  
 MEDLINE  
 PUBMED  
 REFERENCE  
 14  
 AUTHORS  
 TITLE  
 Farsot, C., Cossart, P., Saint-Girons, I. and Cohen, G.N.  
 Nucleotide sequence of *thrc* and of the transcription termination  
 region of the threonine operon in *Escherichia coli* K12  
 Nucleic Acids Res. 11 (21), 7331-7345 (1983)  
 JOURNAL  
 MEDLINE  
 PUBMED  
 REFERENCE  
 15  
 AUTHORS  
 TITLE  
 Bardwell, J.C. and Craig, E.A.  
 Major heat shock gene of *Drosophila* and the *Escherichia coli*  
 heat-inducible *dnaK* gene are homologous  
 Proc. Natl. Acad. Sci. U.S.A. 81 (3), 848-852 (1984)  
 JOURNAL  
 MEDLINE  
 PUBMED  
 REFERENCE  
 16  
 AUTHORS  
 TITLE  
 Innis, M.A., Tokunaga, M., Williams, M.E., Lorranger, J.M., Chang, S.Y.,  
 Chang, S. and Wu, H.C.  
 Nucleotide sequence of the *Escherichia coli* prolipoprotein signal  
 peptidase (*lsp*) gene  
 Proc. Natl. Acad. Sci. U.S.A. 81 (12), 3708-3712 (1984)  
 JOURNAL  
 MEDLINE  
 PUBMED  
 REFERENCE  
 17  
 AUTHORS  
 TITLE  
 Bouvier, J., Patte, J.C. and Stragier, P.  
 Multiple regulatory signals in the control region of the  
*Escherichia coli* *carAB* operon  
 Proc. Natl. Acad. Sci. U.S.A. 81 (13), 4139-4143 (1984)  
 JOURNAL  
 MEDLINE  
 PUBMED  
 REFERENCE  
 18  
 AUTHORS  
 TITLE  
 Bouvier, J., Richard, C., Richard, F., Patte, J.C. and Stragier, P.  
 Nucleotide sequence and expression of the *Escherichia coli* *dapB*  
 gene  
 J. Biol. Chem. 259 (23), 14829-14834 (1984)  
 JOURNAL  
 MEDLINE  
 PUBMED  
 REFERENCE  
 19  
 AUTHORS  
 TITLE  
 Kamio, Y., Lin, C.K., Regue, M. and Wu, H.C.  
 Characterization of the *iles-lsp* operon in *Escherichia coli*.  
 Identification of an open reading frame upstream of the *iles* gene  
 and potential promoter(s) for the *iles-lsp* operon  
 J. Biol. Chem. 260 (9), 5616-5620 (1985)  
 JOURNAL  
 MEDLINE  
 PUBMED  
 REFERENCE  
 20  
 AUTHORS  
 TITLE  
 Friedberg, D., Rosenthal, E.R., Jones, J.W. and Calvo, J.M.  
 Characterization of the 3' end of the leucine operon of *Salmonella*  
 typhimurium  
 Mol. Gen. Genet. 199 (3), 486-494 (1985)  
 JOURNAL  
 MEDLINE  
 PUBMED

REFERENCE  
 21  
 AUTHORS  
 TITLE  
 Consensus sequence for *Escherichia coli* heat shock gene promoters  
 Proc. Natl. Acad. Sci. U.S.A. 82 (9), 2679-2683 (1985)  
 MEDLINE  
 PUBMED  
 REFERENCE  
 22  
 AUTHORS  
 TITLE  
 Chong, P., Hui, I., Loo, T. and Gillam, S.  
 Structural analysis of a new GC-specific insertion element IS186  
 FEBS Lett. 192 (1), 47-52 (1985)  
 JOURNAL  
 MEDLINE  
 PUBMED  
 REFERENCE  
 23  
 AUTHORS  
 TITLE  
 Birnbaum, M.J., Haspel, H.C. and Rosen, O.M.  
 Cloning and characterization of a cDNA encoding the rat brain  
 glucose-transporter protein  
 Proc. Natl. Acad. Sci. U.S.A. 83 (16), 5784-5788 (1986)  
 JOURNAL  
 MEDLINE  
 PUBMED  
 REFERENCE  
 24  
 AUTHORS  
 TITLE  
 Sekiguchi, T., Ortega-Cesena, J., Nosoh, Y., Ohashi, S., Tsuda, K. and  
 Kanaya, S.  
 DNA and amino-acid sequences of 3-isopropylmalate dehydrogenase of  
*Bacillus coagulans*. Comparison with the enzymes of *Saccharomyces*  
*cerevisiae* and *Thermus thermophilus*  
 Biochim. Biophys. Acta 867, 36-44 (1986)  
 JOURNAL  
 MEDLINE  
 PUBMED  
 REFERENCE  
 25  
 AUTHORS  
 TITLE  
 Ohki, M., Tamura, F., Nishimura, S. and Uchida, H.  
 Nucleotide sequence of the *Escherichia coli* *dnaJ* gene and  
 purification of the gene product  
 J. Biol. Chem. 261 (4), 1778-1781 (1986)  
 JOURNAL  
 MEDLINE  
 PUBMED  
 REFERENCE  
 26  
 AUTHORS  
 TITLE  
 Mackie, G.A.  
 Structure of the DNA distal to the gene for ribosomal protein S20  
 in *Escherichia coli* K12: presence of a strong terminator and an IS1  
 element  
 Nucleic Acids Res. 14 (17), 6965-6981 (1986)  
 JOURNAL  
 MEDLINE  
 PUBMED  
 REFERENCE  
 27  
 AUTHORS  
 TITLE  
 Query Match. 1.7%; Score 24; DB 1; Length 111408;  
 Best Local Similarity 100.0%; Pred. NO. 2.6;  
 Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 537 CTCTGAGCTTCCAGCTGGAACCC 560  
 Db 97222 CTCTGAGCTTCCAGCTGGAACCC 97245  
 RESULT 13  
 AC073937  
 LOCUS  
 DEFINITION  
 Mus musculus clone RP23-43M3 strain C57BL6/J, WORKING DRAFT  
 SEQUENCE, 44 unordered pieces.  
 AC073937  
 AC073937.14 GI:14269588  
 HTG: HTGS PHASE1: HTGS DRAFT.  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 248308)  
 REFERENCE  
 1  
 AUTHORS  
 TITLE  
 Montgomery, K.T., Grills, G., Han, J., Lee, E., Long, J., Pomerantz, R.,  
 Toshikhes, I.P., Shim, C., Decker, J., Thomas, E., Perera, A.,  
 Gordon, M., Goltz, J.S. and Kucheralapati, R.  
 High Throughput Mouse Sequencing  
 Unpublished  
 2 (bases 1 to 248308)  
 REFERENCE  
 2  
 AUTHORS  
 TITLE  
 Montgomery, K.T., Grills, G., Han, J., Lee, E., Long, J., Pomerantz, R.,  
 Toshikhes, I.P., Shim, C., Decker, J., Thomas, E., Perera, A.,  
 Gordon, M., Goltz, J.S., Liedtke, W., Friedman, J.M. and  
 Kucheralapati, R.

TITLE  
JOURNAL

Direct Submission  
Submitted (07-JUL-2000) Department of Molecular Genetics, Albert  
Einstein College of Medicine Genome Center, 1300 Morris Park Ave.,  
Bronx, NY 10461, USA  
On May 31, 2001 this sequence version replaced gi:14190542.

## COMMENT

-----Genome Center

Center: Harvard Partners Genome Center

Center Code: HPGC

Web site: <http://www.hpcg.org/Sequence/mouse.html>

Contact: [hpgc@medel.mgh.harvard.edu](mailto:hpgc@medel.mgh.harvard.edu)

-----Summary Statistics

Center project name: AXY

Sequencing vector: pUC18; L08752

Chemistry: Dye-terminator Big Dye; 100%

\*Consensus quality: 231072 at least Q20

\*Consensus quality: 226894 at least Q30

\*Consensus quality: 220122 at least Q40

Estimated insert size: agarose-FP - N/A

\*\*Estimated insert size: 247448 - sum-of-contigs

Quality coverage: agarose-FP - N/A

Quality coverage: 7.4 x in Q20 bases; sum-of-contigs estimation

-----  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 44 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1 45431: contig of 45431 bp in length  
\* 45432 45451: gap of unknown length  
\* 45452 90348: contig of 44897 bp in length  
\* 90349 90368: gap of unknown length  
\* 90369 119208: contig of 28840 bp in length  
\* 119209 119228: gap of unknown length  
\* 119229 134414: contig of 15186 bp in length  
\* 134415 134434: gap of unknown length  
\* 134435 149784: contig of 15350 bp in length  
\* 149785 149804: gap of unknown length  
\* 149805 158841: contig of 9037 bp in length  
\* 158842 158861: gap of unknown length  
\* 158862 166440: contig of 7579 bp in length  
\* 166441 166460: gap of unknown length  
\* 166461 176603: contig of 10143 bp in length  
\* 176604 176623: gap of unknown length  
\* 176624 185346: contig of 8723 bp in length  
\* 185347 185366: gap of unknown length  
\* 185367 192080: contig of 6714 bp in length  
\* 192081 192100: gap of unknown length  
\* 192101 199361: contig of 7261 bp in length  
\* 199362 199381: gap of unknown length  
\* 199382 204102: contig of 4721 bp in length  
\* 204103 204122: gap of unknown length  
\* 204123 207075: contig of 2953 bp in length  
\* 207076 207095: gap of unknown length  
\* 207096 209883: contig of 2788 bp in length  
\* 209884 209903: gap of unknown length  
\* 209904 211730: contig of 1827 bp in length  
\* 211731 211750: gap of unknown length  
\* 211751 213370: contig of 1620 bp in length  
\* 213371 213390: gap of unknown length  
\* 213391 215364: contig of 1974 bp in length  
\* 215365 215384: gap of unknown length  
\* 215385 216600: contig of 1256 bp in length  
\* 216601 216620: gap of unknown length  
\* 216621 217471: contig of 811 bp in length  
\* 217472 220033: contig of 2542 bp in length  
\* 220034 220053: gap of unknown length  
\* 220054 221452: contig of 1399 bp in length  
\* 221453 221472: gap of unknown length  
\* 221473 223540: contig of 2068 bp in length

\* 223541 223560: gap of unknown length  
\* 223561 224311: contig of 751 bp in length  
\* 224312 224331: gap of unknown length  
\* 224332 224943: contig of 612 bp in length  
\* 224944 224963: gap of unknown length  
\* 224964 226352: contig of 1389 bp in length  
\* 226353 226372: gap of unknown length  
\* 226373 227744: contig of 1372 bp in length  
\* 227745 227764: gap of unknown length  
\* 227765 228836: contig of 1072 bp in length  
\* 228837 228856: gap of unknown length  
\* 228857 230267: contig of 1411 bp in length  
\* 230268 230287: gap of unknown length  
\* 230288 231626: contig of 1339 bp in length  
\* 231627 231646: gap of unknown length  
\* 231647 232721: contig of 1075 bp in length  
\* 232722 232741: gap of unknown length  
\* 232742 233649: contig of 908 bp in length  
\* 233650 233669: gap of unknown length  
\* 233670 234512: contig of 843 bp in length  
\* 234513 234532: gap of unknown length  
\* 234533 236145: contig of 1613 bp in length  
\* 236146 236165: gap of unknown length  
\* 236166 237372: contig of 1207 bp in length  
\* 237373 237392: gap of unknown length  
\* 237393 237950: contig of 558 bp in length  
\* 237951 237970: gap of unknown length  
\* 237971 239054: contig of 1084 bp in length  
\* 239055 239074: gap of unknown length  
\* 239075 240538: contig of 1464 bp in length  
\* 240539 240558: gap of unknown length  
\* 240559 241839: contig of 1281 bp in length  
\* 241840 241859: gap of unknown length  
\* 241860 242735: contig of 876 bp in length  
\* 242736 242755: gap of unknown length  
\* 242756 243768: contig of 1013 bp in length  
\* 243769 243788: gap of unknown length  
\* 243789 244721: contig of 933 bp in length  
\* 244722 244741: gap of unknown length  
\* 244742 246391: contig of 1650 bp in length  
\* 246392 246411: gap of unknown length  
\* 246412 247722: contig of 1311 bp in length  
\* 247723 247742: gap of unknown length  
\* 247743 248308: contig of 566 bp in length.  
\* 247743 Location/Qualifiers  
1. 248308  
/organism="Mus musculus"  
/mol\_type="genomic DNA"  
/strain="C57BL6/J"  
/db\_xref="taxon:10090"  
/clone="RP23-439M3"  
/sex="male"  
misc\_feature 1..45431  
/note="assembly\_name:Contig204"  
misc\_feature 45452..90348  
/note="assembly\_name:Contig203"  
misc\_feature 90369..119208  
/note="assembly\_name:Contig202"  
misc\_feature 119229..134414  
/note="assembly\_name:Contig201"  
clone\_end:T7  
vector\_side:left  
134435..149784  
/note="assembly\_name:Contig200"  
misc\_feature 149805..158841  
/note="assembly\_name:Contig199"  
misc\_feature 158862..166440  
/note="assembly\_name:Contig198"  
misc\_feature 166461..176603  
/note="assembly\_name:Contig197"  
misc\_feature 176624..185346  
/note="assembly\_name:Contig196"  
misc\_feature 185367..192080

misc\_feature

/note="assembly\_name:Contig195"

192101. .199361

/note="assembly\_name:Contig194"

199382. .204102

/note="assembly\_name:Contig193"

204123. .207075

/note="assembly\_name:Contig192"

207096. .209883

/note="assembly\_name:Contig191"

209904. .211730

/note="assembly\_name:Contig190

clone\_end:SP6

vector\_side:left"

211751. .213370

/note="assembly\_name:Contig189"

213391. .215364

/note="assembly\_name:Contig188"

215385. .216640

/note="assembly\_name:Contig187"

216661. .217471

/note="assembly\_name:Contig186"

217492. .220033

/note="assembly\_name:Contig185"

220054. .221452

/note="assembly\_name:Contig184"

221473. .223540

/note="assembly\_name:Contig183"

223561. .224311

/note="assembly\_name:Contig182"

224332. .224943

/note="assembly\_name:Contig181"

224964. .226352

/note="assembly\_name:Contig180"

Query Match

Best Local Similarity 1.7%; Score 24; DB 2; Length 248308;

Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY

537 CTGTCGAGCTCCAGCTGGAACC 560

|||||

DB

239530 CTGTCGAGCTCCAGCTGGAACC 239553

RESULT 14

AC020876/c

LOCUS

AC020876 259498 bp DNA linear HTG 29-MAR-2000

DEFINITION

Mus musculus clone RP23-336D4, WORKING DRAFT SEQUENCE, 75 unordered

pieces.

ACCESSION

AC020876

VERSION

AC020876.2 GI:7340302

KEYWORDS

HTG; HTGS\_PHASE1; HTGS\_DRAFT.

SOURCE

Mus musculus (house mouse)

ORGANISM

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE

1 (bases 1 to 259498)

AUTHORS

DOE Joint Genome Institute.

TITLE

Sequencing of Mouse

JOURNAL

Unpublished

REFERENCE

2 (bases 1 to 259498)

AUTHORS

DOE Joint Genome Institute.

TITLE

Direct Submission

JOURNAL

Submitted (10-JAN-2000) Production Sequencing Facility, DOE Joint

Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA

COMMENT

On Mar 29, 2000 this sequence version replaced gi:6686432.

-----Genome Center

Center: Joint Genome Institute

Center Code: JGI

Web site: <http://www.jgi.doe.gov>

-----Summary Statistics

Consensus quality: 178059 bases at least Q40

Consensus quality: 214074 bases at least Q30

Consensus quality: 225088 bases at least Q20

Estimated insert size: 259498; sum-of-contigs estimation

Estimated insert size: 206000; pulse field gel estimation

Quality coverage: 4.28x in Q20 bases; pulse field gel estimation

Quality coverage: 3.40x in Q20 bases; sum-of-contigs estimation

-----

\* NOTE: This is a 'working draft' sequence. It currently

\* consists of 75 contigs. The true order of the pieces

\* is not known and their order in this sequence record is

\* arbitrary. Gaps between the contigs are represented as

\* runs of N, but the exact sizes of the gaps are unknown.

\* This record will be updated with the finished sequence

\* as soon as it is available and the accession number will

\* be preserved.

\* 1 1047: contig of 1047 bp in length

\* gap of unknown length

\* 1048 2408: contig of 1361 bp in length

\* gap of unknown length

\* 2409 3421: contig of 1013 bp in length

\* gap of unknown length

\* 3422 4743: contig of 1322 bp in length

\* gap of unknown length

\* 4744 5783: contig of 1040 bp in length

\* gap of unknown length

\* 5784 6806: contig of 1023 bp in length

\* gap of unknown length

\* 6807 8549: contig of 1743 bp in length

\* gap of unknown length

\* 8550 10037: contig of 1488 bp in length

\* gap of unknown length

\* 10038 11166: contig of 1129 bp in length

\* gap of unknown length

\* 11167 12337: contig of 1171 bp in length

\* gap of unknown length

\* 12338 13402: contig of 1065 bp in length

\* gap of unknown length

\* 13403 14553: contig of 1151 bp in length

\* gap of unknown length

\* 14554 15648: contig of 1095 bp in length

\* gap of unknown length

\* 15649 17323: contig of 1675 bp in length

\* gap of unknown length

\* 17324 18446: contig of 1123 bp in length

\* gap of unknown length

\* 18447 20281: contig of 1835 bp in length

\* gap of unknown length

\* 20282 21651: contig of 1370 bp in length

\* gap of unknown length

\* 21652 22779: contig of 1128 bp in length

\* gap of unknown length

\* 22780 23828: contig of 1049 bp in length

\* gap of unknown length

\* 23829 24969: contig of 1141 bp in length

\* gap of unknown length

\* 24970 26600: contig of 1631 bp in length

\* gap of unknown length

\* 26601 27765: contig of 1165 bp in length

\* gap of unknown length

\* 27766 29202: contig of 1437 bp in length

\* gap of unknown length

\* 29203 30638: contig of 1436 bp in length

\* gap of unknown length

\* 30639 32350: contig of 1712 bp in length

\* gap of unknown length

\* 32351 33814: contig of 1464 bp in length

\* gap of unknown length

\* 33815 34993: contig of 1179 bp in length

\* gap of unknown length

\* 34994 36372: contig of 1379 bp in length

\* gap of unknown length

\* 36373 37711: contig of 1339 bp in length

\* gap of unknown length

\* 37712 38821: contig of 1110 bp in length

\* gap of unknown length

```

* 38822 40400: contig of 1579 bp in length
*      gap of unknown length
* 40401 41429: contig of 1029 bp in length
*      gap of unknown length
* 41430 42543: contig of 1114 bp in length
*      gap of unknown length
* 42544 43957: contig of 1414 bp in length
*      gap of unknown length
* 43958 45245: contig of 1288 bp in length
*      gap of unknown length
* 45246 47013: contig of 1768 bp in length
*      gap of unknown length
* 47014 48155: contig of 1142 bp in length
*      gap of unknown length
* 48156 49158: contig of 1003 bp in length
*      gap of unknown length
* 49159 50170: contig of 1012 bp in length
*      gap of unknown length
* 50171 51392: contig of 1222 bp in length
*      gap of unknown length
* 51393 53265: contig of 1873 bp in length
*      gap of unknown length
* 53266 55279: contig of 2014 bp in length
*      gap of unknown length
* 55280 56908: contig of 1629 bp in length
*      gap of unknown length
* 56909 58527: contig of 1619 bp in length
*      gap of unknown length
* 58528 60093: contig of 1566 bp in length
*      gap of unknown length
* 60094 61444: contig of 1351 bp in length
*      gap of unknown length
* 61445 62672: contig of 1228 bp in length
*      gap of unknown length
* 62673 65545: contig of 2873 bp in length
*      gap of unknown length
* 65546 67540: contig of 1995 bp in length
*      gap of unknown length
* 67541 68584: contig of 1044 bp in length
*      gap of unknown length
* 68585 71572: contig of 2988 bp in length
*      gap of unknown length
* 71573 73312: contig of 1740 bp in length
*      gap of unknown length
* 73313 75668: contig of 2356 bp in length
*      gap of unknown length
* 75669 78367: contig of 2699 bp in length
*      gap of unknown length
* 78368 80669: contig of 2302 bp in length
*      gap of unknown length
* 80670 83363: contig of 2694 bp in length
*      gap of unknown length
* 83364 87042: contig of 3679 bp in length
*      gap of unknown length
* 87043 89727: contig of 2685 bp in length
*      gap of unknown length
* 89728 93097: contig of 3370 bp in length
*      gap of unknown length
* 93098 96142: contig of 3045 bp in length
*      gap of unknown length
* 96143 99546: contig of 3404 bp in length
*      gap of unknown length
* 99547 105084: contig of 5538 bp in length
*      gap of unknown length
* 105085 109494: contig of 4410 bp in length
*      gap of unknown length
* 109495 113825: contig of 4331 bp in length
*      gap of unknown length
* 113826 119314: contig of 5489 bp in length
*      gap of unknown length
* 119315 129224: contig of 9910 bp in length
*      gap of unknown length
* 129225 136674: contig of 7450 bp in length

```

```

* 136675 144863: gap of unknown length
*      contig of 8189 bp in length
* 144864 155259: gap of unknown length
*      contig of 10396 bp in length
* 155260 164645: gap of unknown length
*      contig of 9386 bp in length
* 164646 174742: gap of unknown length
*      contig of 10097 bp in length
* 174743 184830: gap of unknown length
*      contig of 10088 bp in length
* 184831 201876: gap of unknown length
*      contig of 17046 bp in length
* 201877 221979: gap of unknown length
*      contig of 20103 bp in length
* 221980 259498: gap of unknown length
*      contig of 37519 bp in length.

```

# FEATURES

```

source
1..259498
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/clone="Rp23-336D4"

```

```

BASE COUNT 72952 a 58720 c 57559 g 70128 t 139 others
ORIGIN

```

Query Match 1.7%; Score 24; DB 2; Length 259498;

Best Local Similarity 100.0%; Pred. No. 2.3;  
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 537 CTGTCGAGCTTCAGCTGGAACC 560

|||||

Db 50010 CTGTCGAGCTTCAGCTGGAACC 49987

## RESULT 15

```

LOCUS      AP002550      281530 bp      DNA      linear      BCT 07-MAR-2001
DEFINITION Escherichia coli O157:H7 DNA, complete genome, section 1/20.
ACCESSION  AP002550 BA000007
VERSION     AP002550.1 GI:13359456
KEYWORDS
SOURCE      Escherichia coli O157:H7
ORGANISM    Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Escherichia.

```

```

REFERENCE   1 (sites)
AUTHORS     Makino,K., Yokoyama,K., Kubota,Y., Yutsudo,C.H., Kimura,S.,
            Kurokawa,K., Ishii,K., Hattori,M., Tatsuno,I., Abe,H., Iida,T.,
            Yamamoto,K., Ohnishi,M., Hayashi,T., Yasunaga,T., Honda,T.,
            Sasakawa,C. and Shinagawa,H.

```

Complete nucleotide sequence of the prophage VT2-Sakai carrying the verotoxin 2 genes of the enterohemorrhagic Escherichia coli O157:H7 derived from the Sakai outbreak

Genes Genet. Syst. 74 (5), 227-239 (1999)

## JOURNAL

## MEDLINE

## PUBMED

## REFERENCE

## AUTHORS

```

2 (sites)
Ohnishi,M., Murata,T., Nakayama,K., Kuhara,S., Hattori,M.,
Kurokawa,K., Yasunaga,T., Yokoyama,K., Makino,K., Shinagawa,H. and
Hayashi,T.
Comparative analysis of the whole set of rRNA operons between an
enterohemorrhagic Escherichia coli O157:H7 Sakai strain and an
Escherichia coli K-12 strain MG1655
Syst. Appl. Microbiol. 23 (3), 315-324 (2000)
20557356
11108008

```

## 3 (sites)

```

Yokoyama,K., Makino,K., Kubota,Y., Watanabe,M., Kimura,S.,
Yutsudo,C.H., Kurokawa,K., Ishii,K., Hattori,M., Abe,H., Iida,T.,
Yamamoto,K., Hayashi,T., Yasunaga,T., Honda,T., Sasakawa,C. and
Shinagawa,H.
Complete nucleotide sequence of the prophage VT1-Sakai carrying the
Shiga toxin 1 genes of the enterohemorrhagic Escherichia coli

```

## TITLE

gene

CDS

0157:H7 strain derived from the Sakai outbreak  
 Gene 258 (1-2), 127-139 (2000)  
 MEDLINE 20564182  
 PUBMED 11111050  
 4 (sites)  
 AUTHORS Hayashi,T., Makino,K., Ohnishi,M., Kurokawa,K., Ishii,K., Yokoyama,K., Han,C.-G., Ohtsubo,E., Nakayama,K., Murata,T., Tanaka,M., Tobe,T., Iida,T., Takami,H., Honda,T., Sakakawa,C., Ogasawara,N., Yasunaga,T., Kuhara,S., Shiba,T., Hattori,M. and Shinagawa,H.  
 Complete genome sequence of enterohemorrhagic Escherichia coli 0157:H7 and genomic comparison with a laboratory strain K-12  
 DNA Res. 8 (1), 11-22 (2001)  
 21156231  
 11258796  
 5 (bases 1 to 281530)  
 AUTHORS Ohnishi,M., Kurokawa,K., Makino,K., Yasunaga,T., Shinagawa,H. and Hayashi,T.  
 Direct Submission  
 Submitted (26-JUN-2000) Ken Kurokawa, Osaka University, Genome Information Research Center; 3-1, Yamadaoka, Suita, Osaka 565-0871, Japan (E-mail: ken@gen-info.osaka-u.ac.jp,  
 URL: http://www.gen-info.osaka-u.ac.jp/, Tel: 81-6-6879-8365, Fax: 81-6-6879-2047)  
 genome project.

## COMMENT

## FEATURES

source

Location/Qualifiers

1..281530

/organism="Escherichia coli 0157:H7"

/mol\_type="genomic DNA"

/strain="0157:H7"

/sub\_strain="RIMD 0509952"

/db\_xref="taxon:83334"

190..273

/gene="ECs0001"

190..273

/gene="ECs0001"

/note="similar to THRL\_ECOLI gi|1786182 percent similarity 100 in 21 aa, but has 6 additional residues (Conserved in E.coli K-12)"

/codon\_start=1

/evidence=not\_experimental

/transl\_table=11

/product="thr operon leader peptide"

/protein\_id="BAB33424.1"

/db\_xref="GI:13359457"

/translation="MKRISTTITTTITTTITTTITTTGNGAG"

354..2816

/gene="ECs0002"

354..2816

/gene="ECs0002"

/note="similar to THRA\_ECOLI gi|1786183 percent identity 99 in 820 aa (Conserved in E.coli K-12)"

/codon\_start=1

/evidence=not\_experimental

/transl\_table=11

/product="aspartokinase I-homoserine dehydrogenase I"

/protein\_id="BAB33425.1"

/db\_xref="GI:13359458"

/translation="MRVLKFGGTSVANAERFLRVADLTLESNARQGVATVLSAPAKIT  
 NMLVAMIEKTTSGQDALPNISDAERFIELTGLAAQPGFPFLAQKITFVDQEFQAK  
 HVLHGILSGQDPSINAALIRGKEMSIAMAGVLEARGHNVTVIDPEKLLAVGHY  
 LESTVDAESTRIARASIPADHMLVMAAGFTAGNEKGLSVYLGKNGSDYSAVLAACI  
 RADCEIWDVGVYTCDPDPVADRLKMSYQEMELSYFGAGLVLPRTITPTAOF  
 QIPCLIKNTGNPQAPGTLIGASRDELPVKGIISLNNAMESVSGPGMGKMGVGAAR  
 VFAMARSIVSLTQSSEYSISFCVPQSDCVRAERAMQEEYLEKEGLLEPLAV  
 TERLAIISVGDGMLTSGISAKFALARANINIVAIAQSSRSISVNNDDATT  
 GYRVTHMLFNTQDVEYFVIGVGSGGALLEQLKROQSLNKHIDILRVCGVANSKA  
 LITNVHGLNENWQBELAQAKPEPNLGRILRLVKEYHLLNPEYVIVDTSQAVADQYAD  
 FGREGFVVYTNKANTSSMDYHLLRHAAEKSRKFLYDYNVAGLPLVINLONLLN  
 ALDELMKESGILSGLSLYIFGKLDGMSFSATFLAREMGYTEPDPDDLSGMVARK  
 LLILARETGRELELADIEIEPVLPAEFNAGDVAFMANLSQDDLPFAARVAKARDEG  
 KLVRYVGNIDEGGVRVKIAEVDGNDPLFKVKNGENALAFYSHYQPLPLVLRYGAG  
 NDVTAAGVAFADLLRLTSWKLVG"

2818..3750

/gene="ECs0003"

2818..3750

/gene="ECs0003"

/note="similar to THRB\_ECOLI gi|1786184 percent identity 99 in 310 aa (Conserved in E.coli K-12)"

/codon\_start=1

/evidence=not\_experimental

/transl\_table=11

/product="homoserine kinase"

/protein\_id="BAB33426.1"

/db\_xref="GI:13359459"

/translation="MKVYAPASANNMVGFDVLGAATVPDVGALLGDVTVVESAETP  
 SLNLRGFADKLPSREPNIVYQWERFCQELGKQIPVAMTLEKNMPTIGSGIGSSACS  
 VVAALMANHECHGKPLNDTRLALMGELEGRISGIHYDNPVAPCGFLAGMQLMIENDI  
 ISQVPCGFDEWLVLVPGIKVSTAERAILPAQYRRDDCIAGRHLAGETHACYSRQ  
 PELAAKLMKVIAEPYRERLLPGFRQARVAEIGAVASGISGSGPTILFALCDKPDPTA  
 QRVDWLGKNTYLNQEGFVHCRLDTAGARVLE"

3751..5037

/gene="ECs0004"

3751..5037

/gene="ECs0004"

/note="similar to THRC\_ECOLI gi|1786185 percent identity 99 in 428 aa (Conserved in E.coli K-12)"

/codon\_start=1

/evidence=not\_experimental

/transl\_table=11

/product="threonine synthase"

/protein\_id="BAB33427.1"

/db\_xref="GI:13359460"

/translation="MKLYNLKDHNEQVSFAQAVTQGLKNGQLFFPHDLPBFSLTEID  
 EMLKDFVTRSAKILSFIDGIEIPQELIEERVRAFAFAPVAVNVEDSGLELPHGP  
 TLAFGDGGREMAQMLTHIAGDKPVTILTATSGDTGAVAHAFLGLPNVKVILYPRG  
 KISPLQEKLFCTLGNITVAIDGDFACQALVKQAFDDEELKVALGNSANSINISR  
 LLAQICYYFEAVQALPOEARNLVYVPSGNGDITAGLLAKSLGLPKVKTAAATNVN  
 DTVPRFLHDGQWSPKATQATLSNAMDVSNQNNPRVEELFRKKIWLQELKGYAAVDDE  
 TTQOTMRLEKELGTSSEPAVAAYRALDQLNPGEYGLFLGTATHPAKESVESVAILGE  
 TLDLPKELAEARADLPILLSHNLPAFALRKLMMHQ"

5251..5547

/gene="ECs0005"

5251..5547

/gene="ECs0005"

/note="similar to B0005\_ECOLI gi|1786186 percent identity 95 in 98 aa (Conserved in E.coli K-12)"

/codon\_start=1

/evidence=not\_experimental

/transl\_table=11

/product="hypothetical protein"

/protein\_id="BAB33428.1"

/db\_xref="GI:13359461"

/translation="MKKMQSIVLALSLVLVAPMATQAAEITLVPSVKLIQIGDRNNGY  
 YWDGHWDRDHGWWKQHYEWRGNRWHPHGPPPPPPRRHKKAHHDHGHGHPGKHHR"

complement(5700..6476)

/gene="ECs0006"

complement(5700..6476)

/gene="ECs0006"

/note="similar to YAAA\_ECOLI gi|1786187 percent identity 99 in 258 aa (Conserved in E.coli K-12)"

/codon\_start=1

/evidence=not\_experimental

/transl\_table=11

/product="hypothetical protein"

/protein\_id="BAB33429.1"

/db\_xref="GI:13359462"

/translation="MLILISPAKTLDYQSPLTTRYTYTLPDLLNSQQLIHEARKLTTPP  
 QISTRLISDKLAGINIAARFQWQDFTPENARQAILAFKGDVYVGLQAEFTSEDDFD  
 FAQOHLRLSLGYLVRLDLMQYLRLEMGRLENARGLDYQWMDGITITNKLNEALA  
 AQGNVNLASDEYFKSVKPKLNAETIKPVFLDEKNGKFKIISFYAKKARGMLMSRF  
 ILENRLTKPEOLTGFNSGTYFDEASSNGELVFKRYEQR"

complement(6546..7976)

/gene="ECs0007"

complement(6546..7976)

/gene="ECs0007"

/note="similar to YAAJ\_ECOLI gi11786188 percent identity  
99 in 476 aa (Conserved in E.coli K-12)"  
/codon\_start=1  
/evidence=not\_experimental  
/transl\_table=11  
/product="putative inner membrane transport protein"  
/protein\_id="BAB33430.1"  
/db\_xref="GI:13359463"  
/translation="MPDFFSFINSVLWGSVMYLLFGAGCWETFTGFVQFRVIROFG  
KSLKNSITHPOPGGLTSFQSLCTSLAARVSGSNLAGVALAITAGPGGAVFWMWVAEFIG  
MATSEKCSLAQLYKERDVGQPRGGPAWNTMARGLMRWGVLFVFLLIAYGIIFSG  
VQANAVARALUSFDFPPLVTGIILAVFALLIIRGLHGVARLMQGFVPLMAIIWVIT  
SLVICVINIQLPFHVINSIFESAFQWQERAGGAAGYTLQAITNGEQRSMFSNEAGMG  
STPNAAAAASWPPHAAQGIQVQIGIFIDTLVICTASAMLIILAGNGTYMPLEGIG  
LIQKAMRVLMGSGAEFVTLVILFAFSIVANVIVAENNLFFLRLLNNPKAIWCLRIC  
TEATVIGGTLISLPLMWQADIIMACMAITNLTAILLLSPVHTIASDYLRQRKLGV  
PVFDPLRYPDIGRQLSRDWDVYSQ"

Query Match 1.7%; Score 24; DB 1; Length 281530;  
Best Local Similarity 100.0%; Pred. No. 2.3;  
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 537 CTGTCGAGCTCCAGCTGGAACC 560  
Db 102166 CTGTCGAGCTCCAGCTGGAACC 102189  
|||||

Search completed: August 14, 2003, 06:15:42  
Job time : 5311 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 14, 2003, 01:08:54 ; Search time 5305 Seconds  
(without alignments)  
11181.708 Million cell updates/sec

Title: US-09-701-229-1  
Perfect score: 1450  
Sequence: 1 cgtgctgacggcctcgcca.....tggtgacggcgacggcgcac 1450

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2888711 seqs, 2045481386 residues  
Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 100%  
Listing first 45 summaries

Database : GenEmbl.\*

- 1: gb\_ba.\*
- 2: gb\_hgt.\*
- 3: gb\_in.\*
- 4: gb\_on.\*
- 5: gb\_ov.\*
- 6: gb\_pat.\*
- 7: gb\_ph.\*
- 8: gb\_pl.\*
- 9: gb\_pr.\*
- 10: gb\_ro.\*
- 11: gb\_sts.\*
- 12: gb\_sy.\*
- 13: gb\_un.\*
- 14: gb\_vl.\*
- 15: em\_ba.\*
- 16: em\_fun.\*
- 17: em\_hum.\*
- 18: em\_in.\*
- 19: em\_mu.\*
- 20: em\_on.\*
- 21: em\_or.\*
- 22: em\_ov.\*
- 23: em\_pat.\*
- 24: em\_ph.\*
- 25: em\_pl.\*
- 26: em\_ro.\*
- 27: em\_sts.\*
- 28: em\_un.\*
- 29: em\_vl.\*
- 30: em\_htg\_hum.\*
- 31: em\_htg\_inv.\*
- 32: em\_htg\_other.\*
- 33: em\_htg\_mus.\*
- 34: em\_htg\_pln.\*
- 35: em\_htg\_rod.\*
- 36: em\_htg\_mam.\*
- 37: em\_htg\_vrt.\*
- 38: em\_sy.\*
- 39: em\_htgo\_hum.\*
- 40: em\_htgo\_mus.\*
- 41: em\_htgo\_other.\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
C 1	1449	99.9	24000	1	AE004856	AE004856 Pseudomon
2	1396.2	96.3	5280	1	AY008276	AY008276 Pseudomon
3	895.4	61.8	301995	1	AE016779	AE016779 Pseudomon
C 4	827.8	57.1	311600	1	AE016871	AE016871 Pseudomon
5	335.8	23.2	251050	1	AL627265	AL627265 Salmonell
6	335.8	23.2	300169	1	AE016834	AE016834 Salmonell
7	332.6	22.9	22348	1	AE008699	AE008699 Salmonell
8	291.2	20.1	11441	1	AE015046	AE015046 Shigella
9	291.2	20.1	290029	1	AE016978	AE016978 Shigella
10	288	19.9	12791	1	AE005185	AE005185 Escherich
11	288	19.9	281530	1	AP002550	AP002550 Escherich
12	286.4	19.8	300409	1	AE016755	AE016755 Escherich
13	282	19.4	2608	1	ECMUROY	X51584 Escherichia
14	282	19.4	21757	1	AE000118	AE000118 Escherich
15	282	19.4	28277	6	AX191720	AX191720 Sequence
16	282	19.4	28277	15	EC2MIN	X55034 E. coli 2 m
17	282	19.4	111408	1	ECO110K	D10483 Escherichia
18	278.6	19.2	1317	6	AX189058	AX189058 Sequence
19	272.4	18.8	1629	1	ECMURD	X17609 Escherichia
C 20	267.6	18.5	188050	1	AL646072	AL646072 Ralstonia
21	263.8	18.2	248308	2	AC073937	AC073937 Mus muscu
C 22	259.8	17.9	11726	1	AE013965	AE013965 Yersinia
23	253.6	17.5	301727	1	AE016960	AJ414143 Yersinia
C 24	253.6	17.5	301727	1	AE016960	AE016960 Coxella
C 25	212.6	14.7	13071	1	AE015855	AE015855 Shevanel
C 26	212.2	14.6	259498	2	AC020876	AC020876 Mus muscu
27	194	13.4	316050	1	BX321859	BX321859 Nitrosomo
C 28	191.2	13.2	195767	1	NMA722491	AL627358 Neisseria
C 29	185	12.8	22201	1	AE004310	AE004310 Vibrio ch
30	184.8	12.7	298900	1	AP005074	AP005074 Streptomy
31	180.6	12.5	299925	1	AP005045	AP005045 Streptomy
32	180	12.4	10629	1	AE002398	AE002398 Neisseria
33	180	12.4	349980	6	AX044029	AX044029 Sequence
C 34	174.2	12.0	301442	1	AE016798	AE016798 Vibrio vu
C 35	173.8	12.0	37586	1	MTCY270	295388 Mycobacteri
C 36	173.8	12.0	37586	6	AX191745	AX191745 Sequence
C 37	172.2	11.9	20400	1	AE007068	AE007068 Mycobacte
C 38	172.2	11.9	306050	1	BX248341	BX248341 Mycobacte
C 39	160.2	11.0	321250	1	SC093911	AL939111 Streptomy
C 40	155.2	10.7	12401	1	AE005924	AE005924 Caulobact
41	152	10.5	12232	1	AE006048	AE006048 Pasteurel
42	149.8	10.3	11498	1	U32793	U32793 Haemophilus
43	149.8	10.3	11498	6	AX191762	AX191762 Sequence
44	149.8	10.3	110000	6	AR274513	Continuation (12 o
45	149.8	10.3	110000	6	AR274513_12	Continuation (13 o

## ALIGNMENTS

RESULT 1  
AE004856/c  
LOCUS  
DEFINITION Pseudomonas aeruginosa PAO1, section 417 of 529 of the complete genome.  
ACCESSION AE004856 AE004091  
VERSION AE004856.1 GI:9950633  
KEYWORDS  
SOURCE Pseudomonas aeruginosa PAO1  
ORGANISM Pseudomonas aeruginosa PAO1  
Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;  
Pseudomonadaceae; Pseudomonas.  
REFERENCE 1 (bases 1 to 24000)  
AUTHORS Stover,C.K., Pham,X.-Q., T., Erwin,A.L., Mizoguchi,S.D., Warriner,P., Hickey,M.J., Brinkman,F.S.L., Hufnagle,W.O., Kowalik,D.J.,

Lagrou, M., Garber, R.L., Goltry, L., Tolentino, E., Westbrock-Wadman, S., Yuan, Y., Brody, L.L., Coulter, S.N., Folger, K.R., Kas, A., Larbig, K., Lim, R.M., Smith, K.A., Spencer, D.H., Wong, G.K.-S., Wu, Z., and Paulsen, I.T.  
Complete genome sequence of *Pseudomonas aeruginosa* PA01, an opportunistic pathogen  
Nature 406 (6799), 959-964 (2000)  
2043737  
10984043  
2 (bases 1 to 24000)  
Stover, C.K., Pham, X.-Q.T., Erwin, A.L., Mizoguchi, S.D., Warren, P., Hickey, M.J., Brinkman, F.S.L., Huftnagle, W.O., Kowalik, D.J., Lagrou, M., Garber, R.L., Goltry, L., Tolentino, E., Westbrock-Wadman, S., Yuan, Y., Brody, L.L., Coulter, S.N., Folger, K.R., Kas, A., Larbig, K., Lim, R.M., Smith, K.A., Spencer, D.H., Wong, G.K.-S., Wu, Z., Paulsen, I.T., Reizer, J., Saier, M.H., Hancock, R.E.W., Lory, S. and Olson, M.V.  
Direct Submission  
Submitted (16-MAY-2000) Department of Medicine and Genetics, University of Washington Genome Center, University Of Washington, Box 352145, Seattle, WA 98195, USA  
3 (bases 1 to 24000)

*Pseudomonas aeruginosa* Community Annotation Project (PseudoCAP)  
Direct Submission  
Submitted (04-FEB-2003) Department of Molecular Biology and Biochemistry, Simon Fraser University, 8888 University Dr., Burnaby, British Columbia V5A 1S6, Canada

-----  
This represents the February 3, 2003 version of the continually updated, reviewed, *Pseudomonas aeruginosa* PA01 genome annotation, from PseudoCAP (see <http://www.pseudomonas.com> for latest updates and links to alternate annotations). PseudoCAP is coordinated by Fiona S.L. Brinkman (Simon Fraser University, Canada) and Robert E.W. Hancock (University of British Columbia, Canada). We welcome submission through [www.pseudomonas.com](http://www.pseudomonas.com) of any proposed changes.

'Protein name confidence' is used to rate our confidence of the accuracy of the protein name.  
Class 1: Function experimentally demonstrated in *P. aeruginosa*.  
Class 2: Function of highly similar gene experimentally demonstrated in another organism (and gene context consistent in terms of pathways its involved in, if known).  
Class 3: Function proposed based on presence of conserved amino acid motif, structural feature or limited sequence similarity to an experimentally studied gene.  
Class 4: Homologs of previously reported genes of unknown function, or no similarity to any previously reported sequences.

-----  
Location/Qualifiers  
1. .24000  
/organism="Pseudomonas aeruginosa PA01"  
/mol\_type="genomic DNA"  
/strain="PA01"  
/db\_xref="taxon:208964"  
complement(91. .1308)  
/gene="argJ"  
/locus\_tag="PA4402"  
complement(91. .1308)  
/gene="argJ"  
/locus\_tag="PA4402"  
/note="Protein name confidence: Class 2 (high similarity to functionally studied protein); Subcellular localization: Cytoplasmic; Subcellular localization confidence: Class 2; Alt protein name: ornithine transacetylase"  
/codon\_start=1  
/transl\_table=11  
/product="glutamate N-acetyltransferase"  
/protein\_id="AAG07790.1"  
/db\_xref="GI:9950634"  
/translation="NAVGLGPLSTLHPVPGFELGTASGIKRPGKDVVVMRCAGSS VAGVFLNFACPAVTLAKQRFELGEVRYLLTWTGNANAGTGEAGLAAAQTCAKLAEL

AGVAETSVLPYSTGVIGEPVPAKIEAALPAALADLAEDRWAEAAAGIMTTDTLPKGA  
SRQFVGDVTVTIGISKGAMIKPNMATMLGYIATDAKVAQVLODLLRDAANKSFN  
RITIDGDTSTNDCMLIATGRAALPEVTOAGSALFAALQAOVLAVSMEQAIVRDEG  
GATKFTVGVNGGATHQECIDVGYAVAHSPLIKTFALFASPNNGRILAAVGRAGVANL  
DYSKIDVFLGDVCIASRGGRASYTEOAGNAVMAQAEIGIRIELGRGTCSETTWTDDL  
SHEYKINAEIYRT  
complement(1455. .4205)  
/gene="seca"  
/locus\_tag="PA4403"  
complement(1455. .4205)  
/gene="seca"  
/locus\_tag="PA4403"  
/note="Protein name confidence: Class 2 (high similarity to functionally studied protein)"  
/codon\_start=1  
/transl\_table=11  
/product="secretion protein SecA"  
/protein\_id="AAG07791.1"  
/db\_xref="GI:9950635"  
/translation="MFAPLLKFLSGSKNERDVKRMKAQVQAINALEPQMVALSDEQLK  
AKTAEQRYAKGETLDLLPEAFVAVVVRAGKRVGMGRHFDVOLIGMTLHDKGKIAEM  
RTGEGKTLVGLTPVLYNALSGKGVHVTVNDYLARDANMMRPLRYELGLSGVTVPP  
QPEDKRAAYAADITGTTNNEFGDYLRNMAFSLDDKFORELNFVAVDVSILIDE  
ARTPLIISGOAEDSSSELYIKINKLIPRLKRVVEEGKPTTEGHYSIDEKTRVELNE  
QGHQFTEDLSQNLGEGESLYSAHNLSLLTHVYALRALHTFHRYEIVGQDQIL  
LIDHTGRTMPGRRLSEGLHQAPEAKEGPLQAESQTLASTTQNTFRLYKNGMTG  
TADTEAFERQIYGLDVVITPHTRTARKDNFLVLTQEKVAAITDQKQALGR  
PITVGPASIESSEYVSKLQEGAGIEHKLNAKYHEKEETIAQAGAGSVTIATNAG  
RGLDILGNWEVEVALENPTTEQIAQIAEKWKRHQQVIEAGGLHVIASERHESRR  
IDNQLRAGRQDGPSSRFYLSLDNLMRIFASDRVKNFMKALGWSGGAIEHRMVT  
NIAEKARQVEGRNFDIRKOLLEDFVANEQRKVIYHMRNTLLSAEDVGTEIFREE  
TISATNQIHQPQSLQEQMDIEGALYSDFAVRLPIQOMLDEKRLYETRSKIL  
EOTVAAVYEKELAGAEARAFKQMLRLVLDLMDKDLHSLTMDHLRHGLRKYAOKN  
PQOEYRESFTLQELLDSIKRDTIRVLSHVQVRREDPAEEEARLREAEELAKRMQF  
QRAEASMQAVAGEEELPEGPVVPVLEPVNRNEKIGRNEPCGSGKKYKHCHGQ  
LD  
complement(4339. .5262)  
/locus\_tag="PA4404"  
complement(4339. .5262)  
/locus\_tag="PA4404"  
/note="Protein name confidence: Class 4 (homologs of previously reported genes of unknown function, or no similarity to any previously reported sequences)"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="AAG07792.1"  
/db\_xref="GI:9950636"  
/translation="MHIIFLSRHHGAARSVHQLPRLMLAGACVVGASLLAGGFMGAM  
LFRAPAPVAGIEVDAAALQAASAEISSAREDAQMDALSIVHGLGELARLVRLDALGER  
LTELADVDPEDFSLVSGGGVDEPLEGSAYAPPFMATLDELAAALDTRQOQLVYL  
EOLLADKHISOSTYLAGQPVROGYISSPFRVRNVPVSGRLSMHKGIDFAAPAGSDVVA  
VAAGVYVWTSRKNKNGYCNVEVGHADCTTLYAHNQNKLVEVGLDVRKQGVLAQVSTG  
RSTGYVHVHEVMKDGVRMAPQSFARASVE"  
5409. .5804  
/locus\_tag="PA4405"  
5409. .5804  
/locus\_tag="PA4405"  
/note="Protein name confidence: Class 4 (homologs of previously reported genes of unknown function, or no similarity to any previously reported sequences)"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="AAG07793.1"  
/db\_xref="GI:9950637"  
/translation="MLFAEQRICRLOQLLESQLOPAARPYCHVASWRGCGCLLIIVTD  
GHWATRLRYQRRRLQRLQAEFEFANLAKILFKVQPTTSTQRAARSLTUSAVAENL  
RESAEIVGNPRLREALRASAERSQ"  
complement(5865. .6776)  
/gene="lpxC"  
/locus\_tag="PA4406"  
/note="synonyms: envA, asmb"

## TITLE

JOURNAL  
MEDLINE  
PUBMED  
REFERENCE  
AUTHORS

## TITLE

JOURNAL  
AUTHORS  
CONSTIM  
TITLE  
JOURNAL

## COMMENT

FEATURES  
source

gene

CDS

gene



```

RESULT 2
AY008276
LOCUS      5280 bp      DNA      linear      BCT 19-JUL-2001
DEFINITION MurD (murD) genes, complete cds.
ACCESSION AY008276
VERSION    AY008276.2
KEYWORDS   GI:12007715
SOURCE     Pseudomonas aeruginosa
ORGANISM   Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
            Pseudomonadaceae; Pseudomonas.
REFERENCE  1 (bases 1 to 5280)
            Azzolina, B.A., Yuan, X., Anderson, M.S. and El-Sherbeini, M.
            The cell wall and cell division gene cluster in the Mra operon of
            Pseudomonas aeruginosa: cloning, production, and purification of
            active enzymes
            Protein Expr. Purif. 21 (3), 393-400 (2001)
JOURNAL    21178826
MEDLINE    11281713
REFERENCE  2 (bases 1 to 5280)
            El-Sherbeini, M. and Azzolina, B.
            Pseudomonas aeruginosa mray, murD, murE, and murF genes
            Unpublished
JOURNAL
AUTHORS    3 (bases 1 to 5280)
            El-Sherbeini, M. and Azzolina, B.
            Direct Submission
AUTHORS    4 (bases 1 to 5280)
            El-Sherbeini, M. and Azzolina, B.
            Submitted (20-SEP-2000) Biochemistry, Merck Research Laboratories,
            126 E. Lincoln Avenue, Rahway, NJ 07065, USA
JOURNAL
AUTHORS    4 (bases 1 to 5280)
            El-Sherbeini, M. and Azzolina, B.
            Direct Submission
JOURNAL    Submitted (02-JAN-2001) Biochemistry, Merck Research Laboratories,
            126 E. Lincoln Avenue, Rahway, NJ 07065, USA
REMARK     Sequence update by submitter
COMMENT    On Jan 2, 2001 this sequence version replaced gi:10719753.
FEATURES   Location/Qualifiers
            1..5280
                /organism="Pseudomonas aeruginosa"
                /mol_type="genomic DNA"
                /db_xref="taxon:287"
            1..1464
                /gene="murE"
            1..1464
                /gene="murE"
            6..diaminopimelate ligase"
            /codon_start=1
            /transl_table=11
            /product="MurE"
            /protein_id="AAG45235.1"
            /db_xref="GI:12007716"
            /translation="WPMISQLFPOAERDLIRELTDLSHGVRPVDLFLTPVGGHODG
            RAHIDATLKGATAVAYEAGAGELPPSDAPLIAYKGLAAQSAVAGRFYGPSPGLD
            LIGVTNGKTSVQLVAQDLGRCGIVGTGTFYGALESRRHTTDPDLVAQAT
            LATLQAGARVAVYSSHLDQGRVAALFDIYFTNLSDRLDHYHGSMEAYAAKA
            KLFAPDPLRCRYINLDDHGRAGEQDSELITSLTDSALFLYCRASPASPTAR
            RSSIAKOGLLRSLPLGRNENLLAAVAGLILGLVPLDILRTLPLQGPVGRMORL
            GGGKPLVVDYVHTPDALKLEALRPHAAARLLCFCCGDRDRAKRPMLAATER
            LADEVLTDDNPTASAAIADIRKGFNAADKVTFPSRGEAIAHLIASAAVDVVVL
            LAGKHEDIQEDIGVHRHFPDSIEQAERAAWVEVPHA"
            1457..2833
                /gene="murF"
            1457..2833
                /gene="murF"
            /note="recombinant enzyme catalyzed the ATP-dependent
            addition of D-alanine-D-alanine to the
            UDP-N-acetylmuramyl-L-alanine-D-glutamate-m-Dap precursor"
            /codon_start=1
            /transl_table=11
            /product="MurF"
            /protein_id="AAG45236.1"
            /db_xref="GI:12007717"
            /translation="MLEPLRLSQTVALDARLIGEDAVFSVSTDSRAITPGOLFIAL
            SPFRDGHYDLAEVAAKGAVALVEREADAPLPOLLVRDTRAAALGRGLNRRKFTG
            PLAMTSGSKTAVKRWLASILRTQAGDAESVLATRGNLNNDLGVETLLQLAPQHS
            AVTELGASGEIATVELTRPHVAITFAGTAHVGEGFGPEKIVBAKEIGLEGLAAD
            GTAVLNDDKAFDTWKARASGRPLLTSLDRFPQADFPRAADLQDARGCMGFRIGLVAG
            EAQVOLNLGIRHNVAALAAAAHALGVLPLDGIIVAGLQALQPKVGRVAQLTASGLR
            VDDSYNANPASMALAIIDILSGFSGRTVLVLDGMDGLSGWAEQAREVGAYAAKYSVA
            LYAVGLMAHVAQAFGATGRHFADQASLIGALATEQPTTILIKSGSRNMDKVVAAAL
            CQSSBESH"
            2833..3915
                /gene="mray"
            2833..3915
                /gene="mray"
            /note="phospho-N-acetylmuramoyl-pentapeptide translocase"
            /codon_start=1
            /transl_table=11
            /product="Mray"
            /protein_id="AAG22121.1"
            /db_xref="GI:10719754"
            /translation="MLLLAEYLQOFYKGFVQFVQYLTLRGLSVLTALSILWLGPWM
            IFTLOIQIGQAVRNDGPOSHLSKKGKPTPMGALILTAISTLLWADLSNRVYVVL
            VVTLLEAGICGWDDYRKVKIEKNSRGLPSRWKFWQSVFGIGAGAVFLYMTAEPIETTL
            IVMILKSVEIQIGIFIVLVITYFVIGSSNAVNLTDGLDLAIMPYVWVAGALGICFL
            SGNVFAEYLLIPNVPGAGELIVFCAALYVAGLGFELWNTYPAQVPMGVDGVALALGA
            LGTIAIVVRQEIIVLFIMGVFVNETLSVMIOVASFELTGRVFRMAPIHHPFLKGPW
            DPRIVRFRWIIITVILVLIGLATLKLK"
            3929..5275
                /gene="murD"
            3929..5275
                /gene="murD"
            /note="UDP-N-acetylmuramyl-L-alanine:D-glutamate ligase;
            recombinant enzyme catalyzed the ATP-dependent
            addition of D-glutamate to the precursor sugar peptide"
            /codon_start=1
            /transl_table=11
            /product="MurD"
            /protein_id="AAG45237.1"
            /db_xref="GI:12007718"
            /translation="MSLIAADHFRIVVGKSGMSLVRYLARRGLPFAVVDYRENPE
            LATLRAQVQVEARGELDAEFLCSARELYVSPGLSLRTPALVQAQAKVRISGDDIL
            FAREAKAPIVAITGSKNAKSTVTTLGEMAYAADKRVAVGNGLTGTPALDILLADDIELY
            LEISFQLETDRLNNAEVATLVNSDEHMDRDGMADYHLAKHRLFRGARQVYVNRAD
            ALTRPLIADTVPCVFSFGKAFKAFLEEDQKWLAFQKGLKPLVGVSEKIRGAHNY
            SNALAAIGHAVGLPFDAMLGALKAFFSGHRCQWRERQGVSYVDDSKATNVGAAAL
            AAIEGIGADIDGKLVLLAGDGKADFDHDLREPVARFCRAVVLGRDAGLIAQAALGNA
            VPLVRVATLDEAVRQAELAREGDVALLSPACASLDMFKNFEERGLFAKAVEELA"
            817 a 1781 c 1759 g 923 t
BASE COUNT
ORIGIN
Query Match      96.3%; Score 1396.2; DB 1; Length 5280;
Best Local Similarity 99.8%; Pred. No. 6.9e-157;
Matches 1398; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY      2 GTGCTGATCGGCTCGCCACCTTGAAGCTGCGTGTGAGGACGAGAGAGATGAGCCTGAT 61
DB      3880 GTGCTGATCGGCTCGCCACCTTGAAGCTGCGTGTGAGGACGAGAGAGATGAGCCTGAT 9939
QY      62 CGCCTCCGACCACTTCCGCATCGTGTGCGCCTCGGCAAGAGCGGCATGTCCCTGCTGGC 121
DB      3940 CGCCTCCGACCACTTCCGCATCGTGTGCGCCTCGGCAAGAGCGGCATGTCCCTGCTGGC 3999
QY      122 CTACCTGGCGCGCGCGGCTTGCCTTTCCCGCTGCTGATACCCGAGAGAACCCGCCGA 181
DB      4000 CTACCTGGCGCGCGCGGCTTGCCTTTCCCGCTGCTGATACCCGAGAGAACCCGCCGA 4059
QY      182 GCTGGCCACCTTGGTGGCCAGTATCCCGAGGTGGAGTGGTGGCGGCACTCGACGC 241
DB      4060 GCTGGCTACCTTGGTGGCCAGTATCCCGAGGTGGAGGCGCTTGGCGGCACTCGACGC 4119
QY      242 CGAGTTCTCTGCTCGCGCGCGGAACTCTATGTCAGCCCCCGGCTGTGCTCGGCAACCCC 301
DB      4120 CGAGTTCTCTGCTCGCGCGCGGAACTCTATGTCAGCCCCCGGCTGTGCTCGGCAACCCC 4179

```

QY 302 TCGCTGTGTACAGGCGCGCGAAAGGGTGGCATCTCCGGTGACATCGATCTCTTGGC 361  
Db ||||||||||||||||||  
QY 4180 TCGCTGTGTACAGGCGCGCGAAAGGGTGGCATCTCCGGTGACATCGATCTCTTGGC 4239  
Db ||||||||||||||||||  
QY 362 CGCGAGGCGAAGGCCCGATCGTCCCATACCGGTTCCAAAGCGAAGACCGTGCAC 421  
Db ||||||||||||||||||  
QY 4240 CGCGAGGCGAAGGCCCGATCGTCCCATACCGGTTCCAAAGCGAAGACCGTGCAC 4299  
Db ||||||||||||||||||  
QY 422 CACCTGTGTGGCGAAATGGCGTGGCGCGGACAAAGCGTGTGCGCGCGCAACCT 481  
Db ||||||||||||||||||  
QY 4300 CACCTGTGTGGCGAAATGGCGTGGCGCGGACAAAGCGTGTGCGCGCGCAACCT 4359  
Db ||||||||||||||||||  
QY 482 CGGCAACCCCGCGCTCGACCTGCTGGCGCGACGACATCGAGCTGTGCTGGAGCTGTC 541  
Db ||||||||||||||||||  
QY 4360 CGGCAACCCCGCGCTCGACCTGCTGGCGCGACGACATCGAGCTGTGCTGGAGCTGTC 4419  
Db ||||||||||||||||||  
QY 542 GAGCTTCCAGCTGGAAACCTCGGATCGCTCAACCGCGAGGTGGCGACCGTGAACGT 601  
Db ||||||||||||||||||  
QY 4420 GAGCTTCCAGCTGGAAACCTCGGATCGCTCAACCGCGAGGTGGCGACCGTGAACGT 4479  
Db ||||||||||||||||||  
QY 602 CAGCAAGACCATATGATCGTACGACGCGATGCTGACTACCACTGGCCAAAGCACCG 661  
Db ||||||||||||||||||  
QY 4480 CAGCAAGACCATATGATCGTACGACGCGATGCTGACTACCACTGGCCAAAGCACCG 4539  
Db ||||||||||||||||||  
QY 662 GATCTTCCCGGTGCCCGCGAGTGTGATCGCGCGATGCGCTGACCCGACCGCT 721  
Db ||||||||||||||||||  
QY 4540 GATCTTCCCGGTGCCCGCGAGTGTGATCGCGCGATGCGCTGACCCGACCGCT 4599  
Db ||||||||||||||||||  
QY 722 GATCGCGATACCGTGGCGTGTGCTGTGGCGCTGAACAGCGCGACTTCAAGCGTTT 781  
Db ||||||||||||||||||  
QY 4600 GATCGCGATACCGTGGCGTGTGCTGTGGCGCTGAACAGCGCGACTTCAAGCGTTT 4659  
Db ||||||||||||||||||  
QY 782 CGGCTGTATCGAGGAGGCGCAGAGTGGTGGCTTCCAGTTCGACAGCTGCTGCC 841  
Db ||||||||||||||||||  
QY 4660 CGGCTGTATCGAGGAGGCGCAGAGTGGTGGCTTCCAGTTCGACAGCTGCTGCC 4719  
Db ||||||||||||||||||  
QY 842 GGTTCGCAACTGAAGTCCGTGGCGCCACAACTATTCCAAAGCGCTCGCGCGCTGGC 901  
Db ||||||||||||||||||  
QY 4720 GGTTCGCAACTGAAGTCCGTGGCGCCACAACTATTCCAAAGCGCTCGCGCGCTGGC 4779  
Db ||||||||||||||||||  
QY 902 GCTGGCGATCGGCTGGCGTGTGCTGACGCCATGCTGGCGCGCTGAAGCGCTTTTC 961  
Db ||||||||||||||||||  
QY 4780 GCTGGCGATCGGCTGGCGTGTGCTGACGCCATGCTGGCGCGCTGAAGCGCTTTTC 4839  
Db ||||||||||||||||||  
QY 962 CGGCTGTGCTATCGCTGCGAGTGGTACGCGAGCGCGAGGCGTGAAGTACTACGACGA 1021  
Db ||||||||||||||||||  
QY 4840 CGGCTGTGCTATCGCTGCGAGTGGTACGCGAGCGCGAGGCGTGAAGTACTACGACGA 4899  
Db ||||||||||||||||||  
QY 1022 TTCCAAGGCCCAACGTCGCGCGCCCGCTGGCGCGATCGAGGGGCTGGTTCGCGACAT 1081  
Db ||||||||||||||||||  
QY 4900 TTCCAAGGCCCAACGTCGCGCGCCCGCTGGCGCGATCGAGGGGCTGGTTCGCGACAT 4959  
Db ||||||||||||||||||  
QY 1082 CGACGGCAAGCTGTGCTGCTCGCGGGGAGAGCGGCAAGGGCGCGATTCATGACCT 1141  
Db ||||||||||||||||||  
QY 4960 CGACGGCAAGCTGTGCTGCTCGCGGGGAGAGCGGCAAGGGCGCGATTCATGACCT 5019  
Db ||||||||||||||||||  
QY 1142 GCGGAGCGGCTGCGCGCTTCTGCGGGCGGTGTGCTGCTGCGGCTGAGCGCGGCT 1201  
Db ||||||||||||||||||  
QY 5020 GCGGAGCGGCTGCGCGCTTCTGCGGGCGGTGTGCTGCTGCGGCTGAGCGCGGCT 5079  
Db ||||||||||||||||||  
QY 1202 GATTGCCAGGCACTGGGCAACGCGGTACCGCTGGTGGCGCTGCGCAAGCGCTGGAGCAAGC 1261  
Db ||||||||||||||||||  
QY 5080 GATTGCCAGGCACTGGGCAACGCGGTACCGCTGGTGGCGCTGCGCAAGCGCTGGAGCAAGC 5139  
Db ||||||||||||||||||  
QY 1262 AGTCGGGAGGCGCGGAGCTGGCGCGGAGAGCGCATGCGGTGCTGTCGCCGCGCTG 1321  
Db ||||||||||||||||||  
QY 5140 AGTCGGGAGGCGCGGAGCTGGCGCGGAGAGCGCATGCGGTGCTGTCGCCGCGCTG 5199  
Db ||||||||||||||||||  
QY 1322 CGCGAGCTGGACATGTTCAAGAACTTCGAAGAAAGCGGACCGCTGTCGCGCAAGCGCT 1381  
Db ||||||||||||||||||  
QY 5200 CGCGAGCTGGACATGTTCAAGAACTTCGAAGAAAGCGGACCGCTGTCGCGCAAGCGCT 5259  
Db ||||||||||||||||||

QY 1382 AGAGGAGTACGCTGATGCTG 1402  
Db ||||||||||||||||||  
QY 5260 AGAGGAGTACGCTGATGCTG 5280  
Db ||||||||||||||||||  
RESULT 3  
AE016779 301995 bp DNA linear BCT 12-DEC-2002  
Pseudomonas putida KT2440 section 6 of 21 of the complete genome.  
LOCUS AE016779 AE015451  
DEFINITION AE016779.1 GI:26557023  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
Pseudomonas putida KT2440  
Pseudomonas putida KT2440  
Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;  
Pseudomonadaceae; Pseudomonas.  
REFERENCE  
1 (bases 1 to 301995)  
Nelson,K., Paulsen,I., Weinel,C., Dodson,R., Hilbert,H., Fouts,D.,  
Gill,S., Pop,M., Martins Dos Santos,V., Holmes,M., Brinkac,L.,  
Beanan,M., DeBoy,R., Daugherty,S., Kolonay,J., Madupu,R.,  
Nelson,W., White,O., Peterson,J., Khouri,H., Hance,I., Lee,P.,  
Holtzapple,E., Scanlan,D., Tran,K., Moazzez,A., Utterback,T.,  
Rizzo,M., Lee,K., Kosack,D., Moesti,D., Wedler,H., Lauber,J.,  
Hohseil,J., Straetz,M., Helm,S., Kiewitz,C., Eisen,J., Timmis,K.,  
Duesterhoft,A., Tummier,B. and Fraser,C.  
Complete genome sequence and comparative analysis of the  
metabolically versatile Pseudomonas putida KT2440  
Environ. Microbiol. 4 (12), 799-808 (2002)  
2 (bases 1 to 301995)  
Nelson,K., Paulsen,I., Weinel,C., Dodson,R., Hilbert,H., Fouts,D.,  
Gill,S., Pop,M., Martins Dos Santos,V., Holmes,M., Brinkac,L.,  
Beanan,M., DeBoy,R., Daugherty,S., Kolonay,J., Madupu,R.,  
Nelson,W., White,O., Peterson,J., Khouri,H., Hance,I., Lee,P.,  
Holtzapple,E., Scanlan,D., Tran,K., Moazzez,A., Utterback,T.,  
Rizzo,M., Lee,K., Kosack,D., Moesti,D., Wedler,H., Lauber,J.,  
Hohseil,J., Straetz,M., Helm,S., Kiewitz,C., Eisen,J., Timmis,K.,  
Duesterhoft,A., Tummier,B. and Fraser,C.  
Direct Submission  
Submitted (05-NOV-2002) The Institute for Genomic Research, 9712  
Medical Center Dr, Rockville, MD 20850, USA  
FEATURES  
Location/Qualifiers  
1..301995  
/organism="Pseudomonas putida KT2440"  
/mol\_type="genomic DNA"  
/strain="KT2440"  
/db\_xref="taxon:160488"  
89..712  
/gene="sspa"  
/locus\_tag="pp1320"  
89..712  
/gene="sspa"  
/locus\_tag="pp1320"  
/note="similar to GP:9864180; identified by sequence  
similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="stringent starvation protein A"  
/protein\_id="AAN66944.1"  
/db\_xref="GI:2498277"  
/translation="MGATNRLACYSDPADHYSHRVLRLAEGVSVQILDVDPALHLP  
KLAENYPYGSVPTLVDRDLALTEYVWMEYLEEYPPHPLMPVYPVARGNSRLMHRI  
QRDWCLADATVLDPRSSEAAATEARKALRESLTGVSPLFSEFACFMSRDSQSLVDCCLL  
PTLWRLPVLGIELPQAKPLLDYMERQFAREFQASLSVSVERMKL"  
693..1154  
/gene="sspb"  
/locus\_tag="pp1321"  
693..1154  
/gene="sspb"  
/locus\_tag="pp1321"  
/note="similar to GB:S50157, GB:X60221, SP:P24539, and  
PFD:509291; identified by sequence similarity; putative"  
/codon\_start=1  
/transl\_table=11









Query Match	57.1%; Score 827.8; DB 1; Length 311600;
Best Local Similarity	73.6%; Pred. No. 6.5e-90;
Matches 1052; Conservative	0; Mismatches 370; Indels 0; Gaps 0;

QY	21	CTTTGAAGCTGCTGTTGAGGACGAAGACGATGAGCCTGATCGCTCCGACCACTTCGCG	80
DB	291882	CTTTGCCACACTGAAACTGAGTAGACGAGTGTCCCTGATGCTTTCCGACCGCTTCGCG	291823
QY	81	ATCGTTGTGCGGCTCGGCAAGAGCGCATGTCCTGTTGTCGCTACCTGGCGCGCGCGCG	140
DB	291822	ATTGTGTGGGCTTGGCAAGAGTGGCATGTCCCTGTTGCTTCTCGCGCAACCGCGCG	291763
QY	141	TTGCGCTTTGCGCGTGGTGFATACCGAGACGACCGCGGAGCTGGCCACCTCGGTGCC	200
DB	291762	GTGTCTTTGCTTGGCGGACACGCGGAAATCTCTGAGCTGGCGACTTTCGCTGCT	291703
QY	201	CAGTATCCGACGTTGAACTGCTTGGCGGCAACTCGACGCGAGTTCCTCTGTCGCGCC	260
DB	291702	GACTATCCTTCAGTGGAGTGGGTGGGAACTGGAGCTGGACTTCTCTGCGGTGCC	291643
QY	261	CGGGAACCTATGTAGCGCCCGCTTGTGCTGTGCGCACCTCTGCGCTGGTACAGCGCGCC	320
DB	291642	GACGAGCTTTATGTACGCCCTGTGTGCGCCTTGGCCACACCGCGCTTTCAGCAGCGCAT	291583
QY	321	GGGAAGGCGTGGCATCTCCGCTGACATCGATCTTTCGCGCGAGCGGCAAGCGCCCG	380
DB	291582	GCACGGCGCGCAAGCTGTCCCGGCGACATTCGAACCTGTTTGGCCGTTACGCCAAGGACCG	291523
QY	381	ATCGTGGCATCACCGGTTTCCAAACGCGAAGACACCGTGCACACCTGTGGTGGCGAAATG	440
DB	291522	GTGATTGCGATTACCGGTTTCCAAACGCGAAGACACCGTGACACGCTGTGCGGAGATG	291463
QY	441	CGGTTGGCGCGGACAGGCTGTGCGCTGCGGCGCACTTCGGCACCCCGCGGCTCGAC	500
DB	291462	CGCGCAGCTGTCCGCGCAAGGCTTGTCTGTGGCGGCAACCTTGGGCAACCCCGCGCTGAC	291403
QY	501	CTGCTGGCGGACGATCGAGCTGTACGTTGTGGAGCTGTGAGCTTCCAGCTTGGAAACC	560
DB	291402	CTGCTCAGCGATGATCTCAGTGTGTACGTGATGGAATGTCCAGTTTCAGCTCGAAACC	291343
QY	561	TGGATTCGCTCAACCGCGAGTGGCGACCGTGTGTAACGTCAGCGAAGACCATATGAT	620
DB	291342	ACTGATCAGCTCAATGCCGAAGTGGCGACCGTGTGAACATCAGCGAAGACCATATGAC	291283
QY	621	CGCTACGAGCGATGCTGCTACACACTGCGCAAGCAGCAGGATCTTCGCGGTGCGCGC	680
DB	291282	CGTTATTCGCTGTCCGCGCTTATCACCTTGCACAGCAGCGATTTTCCGTGGCGCGCG	291223
QY	681	CAGTGTGGTGAATCGCGCGGATGCCCTGACCGACCGCTGATCGCGATACCTGTGCGG	740
DB	291222	CAAGTAGTGTCAATCGGCGAGCAGCGCTGTCCGCTGCTGATCGCGAAGGCTGTGCT	291163
QY	741	TGCTGTGCTGTGGCTGAAACGCGGACTTCAAGGCTTTGGGCGCTGATCGAGGAAGAC	800
DB	291162	TGCTGAGCTTGTGGCTCAACAGCGCTGACTTCCACGCGCTTGGTCTCGGTGAAGAGA	291103
QY	801	GGCCAGAAGTGGCTGCGCTTCAGTTTCGACAGCTGCTGCGGTTGGGCAACTGAAGATC	860
DB	291102	GGCGAAATTCCTGGCATTCAGTTTCGAGAACCTTGATCGCGGTGAGCGAGCTGAAGATT	291043
QY	861	CGTGGCGCCCAACACTATTCCAAACGCGTTCGCGCGCTGGCGCTGGGCGCATCGGTGCG	920
DB	291042	CGCGGCGCACAACTACAGCTTAACGACTGGCAGCCCTGGCGCTGGGCGATCGGTGGCG	290983
QY	921	CTGCCCTTCGACGCCATGCTCGCGCGCTGAAGGCGTTTTCGCGCTGGCTCATCGCTGC	980
DB	290982	CTGCCCTTCGACGCCATGCTTCCAGCCTTCGCTGAGTTTCCGCGCTTTCGAGCACCGCTGC	290923
QY	981	CAGTGGGTACGCGAGCGGCGAGGCGTACTACTAGCAGGATTCGAAGGCCACCAAGCTC	1040
DB	290922	CAATGGCTCGTGAAGCGCAATGGCTGGGACTACTACAAGGATTTCCAAGGCCACCAAGCTG	290863

1041	GGCGCCCTTCGCGGCGATGAGGCGCTGGGTGCGCGATCGACATCGACGCGCGAGCTGCTG	1100
290862	GGTCTGCACTGGCGCGCTATCGAAGCGCTGGGCTCGGATATCGCGCGCAAGCTGGTGTG	290803
1101	CTCGCGCGGAGACGCGAAGCGCGGATTTCCATGCTGGCGAGCGGTCGCGCGC	1160
290802	ATCGCGCGTGGCATGCGCAAGGTGCGGACTTCAGCGGCTGGCTGCCCGTGGGTAAA	290743
1161	TTCTGCGCGGCGGTGCTGCTGTTGCGCGTGTGACGCGCGGCTGATGCCAGGCACTGGGC	1220
290742	TACTGCGCGCTCGGCTGTTGCTGGGTGCGGATGCGGAGCTGATTTGCCAGCGCTGGT	290683
1221	AAGCGGTACCGTGGTGGCGCTCGCAACGCTGGAGCAAGCACTCGCGGAGCGCGCGAG	1280
290682	GATGCGGTGCGCTGATTCGTGTCGACAGGTTTCAGCGGCTGCCGAGCGCGCGAG	290623
1281	CTGCGCGCGAAGCGGATGCGGTGCTGTTGCTCGCGCGCTTCGCGGAGCTGGACATGTC	1340
290622	CTGCTCAGCGCGCATGCGGTGCTGTTTCCCTGTCATGCGCCAGTCTGGACATGTC	290563
1341	AGAAGCTTCGAAGACGCGGACGCTGTTGCGCAAGCGCTAGAGAGCTAGGCTGATGC	1400
290562	AGAATTTATGAAGCGCGGACGCTGTTGTCGCGAGCGAGTGGAGTGTGTCATGATCT	290503
1401	TGCGGTGTTGCGCCCTTCGCGCTGCGCTGTTGAGCGCGGACGCACTC	1450
290502	TCGCTGTCATCAAGCCCTTATCCATTCGCACTGATCAGCGCGCGGTATC	290453

RESULT 5	AL627265	251050 bp	DNA	linear	BCT 06-JUN-2002
LOCUS	Salmonella enterica serovar Typhi (Salmonella typhi) strain CT18,				
DEFINITION	complete chromosome, segment 1/20.				
ACCESSION	AL627265	AL513382			
VERSION	AL627265.1	GI:16501283			
KEYWORDS	Salmonella enterica subsp. enterica serovar Typhi				
SOURCE	Salmonella enterica subsp. enterica serovar Typhi				
ORGANISM	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;				
REFERENCE	1 (bases 1 to 251050)				
AUTHORS	Parkhill, J., Dougan, G., James, K.D., Thomson, N.R., Pickard, D.,				
	Wain, J., Churcher, C., Mungall, K.L., Bentley, S.D., Holden, M.T.G.,				
	Sebahia, M., Baker, S., Basham, D., Brooks, K., Chillingworth, T.,				
	Conerton, P., Cronin, A., Davis, P., Davies, R.M., Dowd, L., White, N.,				
	Farrar, J., Feltwell, T., Hamlin, N., Haque, A., Hien, T.F., Holroyd, S.,				
	Jagels, K., Krogh, A., Larsen, T.S., Leather, S., Moule, S., O'Gaora, P.,				
	Perry, C., Quail, M., Rutherford, K., Simmonds, M., Skelton, J.,				
	Stevens, K., Whitehead, S. and Barrell, B.G.				

```

gene
190. .255
/genes="STY0001"
/notes="synonym: thrL"
190. .255
/genes="STY0001"
/notes="Orthologue of E. coli thrL (LPT_ECOLI); Fasta hit
to LPT_ECOLI (21 aa), 86% identity in 21 aa overlap"
/codon_start=1
/transl_table=11
/product="thr operon leader peptide"
/protein_id="CAD01154.1"
/db_xref="GI:16501284"
/db_xref="SPTREMBL:O8XG12"
/translating="MNRISTTTTITTTGNGAG"
337. .2799
/genes="STY0002"
/notes="synonym: thrA"
337. .2799
/genes="STY0002"
/notes="Orthologue of E. coli thrA (AKH_ECOLI); Fasta hit
to AKH_ECOLI (820 aa), 94% identity in 820 aa overlap"
/codon_start=1
/transl_table=11
/product="aspartokinase I/homoserine dehydrogenase I"
/protein_id="CAD01155.1"
/db_xref="GI:16501285"
/db_xref="SPTREMBL:O8Z9R7"
/translating="NRVLKFGTGSVANAERFLRVADILLESNSRQGOVATVLSAPAKIT
NLHVMIEKTIGQDALPISDAERFLSLLAGLSAQGFPLARLKNVVEEFAQIK
HVLGSLGQCPSINAALICRGENMSTAIWAGLLEARGHRVTIDPVEKLVAGHY
LSTVDIAESTRIASQIPADHMLMAGTAGNEKGEVLVLRNGSDYSAVALAACL
RADCEIVTDVGVTCDPQRPDARLLKMSYQAMLSYFCAKVLHPRTTTPAQL
QIPCLIKTGNPQAGTLLIGSSDDNLPVIGISLNNMAMFVSQGMKGIMGAAR
VFAAMSRAGISVVLITQSSEYISFCVQPCARARMAQDEFVLEKLEGLLEPLAV
TERLAIISVVGDMETLGCISAKFAALANINIVAIAQSSERSISVYVNDGATT
GYRVHQLMFTDQVIEVFVIGVGGVGGALLEGQLKQQLNKKHIDLRVCVANSKA
LUTNHGLNDWQLEAQAANPFLNRLIRLVKEYLHLLNPVIVDTSSQVADQYAD
FLREGFVYTPNKKATSMDDYHQLRFAAQSRRKFLDYTNVAGLPIVNIQNLN
AGDELQKFSGLSGLSFFIGLECMSLSQALAREMGYTEPDPRDLSQMDVARK
LILARETGRELESDIVIEPLVDFEASGVDTAFMAHLPOLDDAFAARVAKARDEG
KVLRVVGNTEEDGVCVKTAEDVNDPLFKVNGENALAFYSHYQPLPLVLRGYGAG
NDVTAAGVPADLLRLSLWKLGV"
337. .1101
/genes="STY0002"
/notes="pfam match to entry PF00696 aakinese, Amino acid
kinase family, score 207.70, E-value 1.7e-58"
343. .369
/genes="STY0002"
/notes="PS00324 Aspartokinase signature"
1819. .2781
/genes="STY0002"
/notes="pfam match to entry PF00742 Homoserine dh,
Homoserine dehydrogenase, score 535.60, E-value 3.4e-157"
2314. .2382
/genes="STY0002"
/notes="PS01042 Homoserine dehydrogenase signature"
2801. .3730
/genes="STY0003"
/notes="synonym: thrB"
2801. .3730
/genes="STY0003"
/notes="Orthologue of E. coli thrB (KHSE_ECOLI); Fasta hit
to KHSE_ECOLI (310 aa), 94% identity in 308 aa overlap"
/codon_start=1
/transl_table=11
/product="homoserine kinase"
/protein_id="CAD01156.1"
/db_xref="GI:16501286"
/db_xref="SWISS-PROT:O8XGP5"
/translating="MVKYVAPASSANMSVGFVGLAAVTPVDGTLIGDWSVEAADHF
VHVLGRFADKLPPREINIVYQCWERFCQALGKIPVAMTLEKNMIGSGLGSSACS
RVAAVAMNHCGRPLNDRLLALMELEGRISGSIHVDNVAPCPFLGGMGLMIENGII
ISQVPGDEMLWLVAYPGIKVSTAEARAILPAQYRRQDCLAHGRHLAGFIHACYRSR

```

```

POLAALMKDVIAEYRARKLPGFSQARQAVSEIGALASIGSGPTLFCALCKPETA
QRVADWLSKHYLQNGEFVHICRLDTAGARVVG"
3005. .3196
/genes="STY0003"
/notes="pfam match to entry PF00288 GHMP_kinases, GHMP
kinases putative ATP-binding proteins, score 79.20,
E-value 5.7e-21"
3068. .3103
/genes="STY0003"
/notes="PS00627 GHMP kinases putative ATP-binding domain"
3734. .5020
/genes="STY0004"
/notes="synonym: thrC"
3734. .5020
/genes="STY0004"
/notes="Orthologue of E. coli thrC (THRC_ECOLI); Fasta hit
to THRC_ECOLI (428 aa), 93% identity in 428 aa overlap"
/codon_start=1
/transl_table=11
/product="threonine synthase"
/protein_id="CAD01157.1"
/db_xref="GI:16501287"
/db_xref="SPTREMBL:O8Z9R6"
/translating="MKLYNLKDHNEQVSFAQAVTQGLGKQQLFFPHDLPFSLFEID
ENLQDFVRSARILSFAIGDEIPQOILEERVAAFAFPAPVAQVEDVGCLELPHGP
TLAFKDFGRFNAQMLTHISGDKPVTILTATSGDTCAVAHAFYGENVYVILYPRG
KISPLQKLFCTLGGNIETVADGDFACQALVKQAFDDDEELKTALGNSANSINISR
LQAQICFEVAQVLPQAGARNOLVSPSGNGDLTAGLAKSLGLPVKRFRTAANNIN
DTVPFRLHDGKAPKATQNTLSNAMDVSQPNWPRVEELFRKRLWELFGLVAAVDDTT
TYQPMRELUKAGYISEPHAAVAYRALDQNLNPGPEYGLFGLTAHFAKFKESVESILGE
TIALPEALAEARADLLPILLSHLLPADFAALRKLMMTRQ"
3962. .4867
/genes="STY0004"
/notes="pfam match to entry PF00291 PALP,
Pyridoxal-phosphate dependent enzyme, score 104.50,
E-value 2e-27"
4022. .4066
/genes="STY0004"
/notes="PS00165 Serine/threonine dehydratases
pyridoxal-phosphate attachment site"
complement(5114. .5887)
/genes="STY0005"
/notes="synonym: yaaA"
complement(5114. .5887)
/genes="STY0005"
/notes="Orthologue of E. coli yaaA (YAAA_ECOLI); Fasta hit
to YAAA_ECOLI (258 aa), 86% identity in 257 aa overlap"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="CAD01158.1"
/db_xref="GI:16501288"
/db_xref="SPTREMBL:O8Z9R5"
/translating="MLIISPAKTLDYQSPLATTRYTOPELLDHSQQLIQOARQLSAP
QISRLMGISDLADLNATRFHDWQPHFTPDNARQAILAFKGDVYTGLEQTFNDADF
FAQQHLRLMSGLYGLVRLDLMQPYRLEMRGRKLDLYQFQGDITDKLNALE
AQGRVYVNLASSEYFKSPKLDALIKPVFLDKRNGKFWVSFYAKKARGLSRF
IIEENRLTKPEQLTAFDREGYFDEETSTQDELVFKRYEQ"
complement(5966. .7396)
/genes="STY0006"
/notes="synonym: yaaJ"
complement(5966. .7396)
/genes="STY0006"
/notes="Similar to Bacillus subtilis amino acid carrier
protein alst SW:ALST_BACSU (Q45068; P40743) fasta
scores: E(): 0, 45.3% id in 477 aa, and to Alteromonas
haloplanktis na(+)-linked D-alanine glycine permease DGA
SW:DGA_ALTHA (P30144) fasta scores: E(): 0, 36.5% id in
315 aa
Orthologue of E. coli yaaJ (YAAJ_ECOLI); Fasta hit to
YAAJ_ECOLI (476 aa), 76% identity in 476 aa overlap"
/codon_start=1
/transl_table=11

```

		/product="putative amino-acid transport protein"	
		/protein_id="CAD01159.1"	
		/db_xref="GI:16501289"	
		/db_xref="SPTREMBL:Q8Z9R4"	
		/translation="MPEFFSTINELMGSMVYLLLAGAGCWFTWRTGFIQPRYTRQFS RSLKGSUSPQGGTSTLAARIGSGNLGVALIAAAGPGPAGVFMVWVAIIG MATFAECSLAQLYKERDPTQFRGPANMARGIMRWGVVFLPLVAYGLIFNS VQANVASRALHFAFNPLISGLIALAFCALLIIRIGIKVARLWMLPLIALLLWVAG SVFTCLWHIEOMPVGVIASIVKSAQWDEAAAGAAGYTLTOAITSGFORMFSNEAGMG STPNAAATSYPPHPVAQGVQMGVFEVDPIILICTASAMILLIAGNHASHSSTEGIQ LLQHNWLSLTGEWASFVALIVILFAPSSIVANVYIAENNLFFLRHNAKAIWLLRLA TLGMVAGTLISFLIWLQADMIMACMAITNLITALLSPVYITLAGDYLKORKLGV	
Query Match		23.2%; Score 335.8; DB 1; Length 251050;	
Best Local Similarity		56.2%; Pred. No. 1.5e-31;	
Matches 750; Conservative		0; Mismatches 552; Indels 33; Gaps 5;	
QY	69	GACCACTCCGATCGTGTGGCTCGCGAAGACGGCATGTCCTGGTGGCTACCTG	128
DB	148829	GACAAAAACGTGCTATTCCTGGCTTAACCGACTCTCTTGGCTGACTTTTTT	148888
QY	129	GCGGCGCCGGCTTCCCTTTCCGCGTGTGCTATACCCGAGAGAACCCGCGAGCTGGCC	188
DB	148889	CTCGCCCGGGGTGACCCGGGGTGATGATACCTCCGTGACCCCGCGGGTCTGGAT	148948
QY	189	ACCTCGCTGCCAGTATCCGAGGTGGAAGTGCCTTGCAGCACTCGACGCCAGTTC	248
DB	148949	AAGCTGCCG-----CAAGAGGTTGAGCGTCACTGTTGGCGGCTGAACGACGAGTGG	148999
QY	249	CTCTGCTCGCCCGCACTATGATCGACCCCGCGCTTGTGCTGCGGACCCCTCGCGTG	308
DB	149000	CTCTTAGCGGGGGATTTAATCGTCCGACCCCTGTGTATGCTTGGCGCATCCCTCTCTT	149059
QY	309	GTACAGGCGCGCGGAAAGGCGTGCATCTCCGCTGACATCGATCTCTTCCCGCGCAG	368
DB	149060	AGCGTGGCGCAGCGCTGGAGTGGAATTTGCGCGATATCGAACTGTTTTCGCGCGAA	149119
QY	369	GCGAAGCCCGCATCGTCCCATCACCGGTTCCAAACGGAAGACCGCTGACCAACCCCTG	428
DB	149120	GCGAAGCGCGGATTTGGCCATCACCGGCTCGAAGCGCAAAAGCACCGTGAACCACTTA	149179
QY	429	GTGGCGGAANTGGCGGTGGCGCGCAAGCGTGTGCGCGTGGCGGCAACCTCGGCACC	488
DB	149180	TTTGGCGAGATGGGGAAGCGGCGCGCGCTCAATTTGCGCGTGGCGCAATATCGCTCTT	149239
QY	489	CGGCGCTCGACTCTCTGGCGCAGCATCGAGCTGTACGTGTTGCGAGCTGTCGAGCTTC	548
DB	149240	CGGCGCTGTGCTGCTGGATGCCGATCGGAAATTTGACGTGCTGGAATATCCAGTTTC	149299
QY	549	CAGCTGGAAACCTCGCATCGCTCAACCGCAGGTGGCGACCGTGTCTGAAGCTCAGCGAA	608
DB	149300	CAATTGGAAACTACCTCAAGTTTGAAGCGCGCGCAACCGTGTCTCAAGCTCACTGAA	149359
QY	609	GACCATATGGATCGCTA---CGACGGCATGGCTGACTACCACTGGCGCAAGCACCGGATC	665
DB	149360	GATCATATGGACCGGTATCCCTTTGTTGCAACAGTACCGGCGCGCGCAAACTGCGCGTC	149419
QY	666	FTCCGCGGTGCGCGCAGCTGTGTGATCGCGCGGATGCCCTACCCGACCGCTGATC	725
DB	149420	TACGAGAGGCGAAAGTGTGGCTAGTGAATCCGATGACGGTGTGACGATCGCGGTACGC	149479
QY	726	GCCGATACCGTCCGCTGTGCTGTTCGCGCTGAAACAGCGCGACTTCAAGCTTTCCGC	785
DB	149480	GGGGCCGATGAGCGCTGGCTAGCTTTGGCGTCAATATGGGTGATATCA-----C	149530
QY	786	CTGATCGAGGAGCGGCGAGAGTGGCTGGCTTCCAGTTCGACAAGCTGCTCGCGGTT	845
DB	149531	CTTAATCTGACGAGCGGAAACCTTGGCTACGAGTCAAAAGGTGAGAAGGTGCTGAACGTG	149590
QY	846	GCGCAACTGAGATCCGTGGCGCCCAACAATTTCCACGCGCTCGCGCGCTGCGCGTG	905
DB	149591	AAAGAGATGAAGCTTTCCGGTTCAGCATAACTATACCAATGCTTAGCGCGCGCTGGCGTG	149650

QY	906	GGCCATCGCGTGGCTCCCTTCGACGCCATGCTCGGCGCGCTGAAGCGGCTTTCCCGC	965
DB	149651	GCGGATGCGGTAGGTCTGCGCGCGCCAGCAGTTTGAAGCGGTTGACGACATTTACCGGT	149710
QY	966	CTGGCTCATCGCTGCCATGGGTACGAGCGGCGAGGCGGTGAGTACTACGACGATCC	1025
DB	149711	CTGGCGCATCGCTTCCAGCTGGCGTTGGAGCATTAACGCGTTCGTTGGATCAACGACTCG	149770
QY	1026	AAGGCCACCAACGTCGCGCGCCCTGCGCGCGATCGAGGGGCTGGGTGCGGACATCGAC	1085
DB	149771	AAAGCGACCAATGTCGCGCAGTACCAAGCGGCGCTAAACGGTTTGCATG-----TGGAC	149824
QY	1086	GCAAGCTGGTGTCTGCTGCGCGGAGAGCGCAAGGCGCGGATTTCCATGACCTCGCG	1145
DB	149825	GGTACGCTGCATCTGCTGCTGGGCGGCGAGCGTAAGTCGGCAGACTTTTCTCCGCTGGCG	149884
QY	1146	GAGCGGCTCGCGGCTTCTCGCGGCGGTGCTACTGCTTGGCGGCTGACCGCGGCTGATT	1205
DB	149885	CGCTATCTGACCGCGGATGCTATCCGCTGTATTGCTTTGGGCGGATGGCGCGAGCTT	149944
QY	1206	GCCAGGCACTGGGCAACGCGGTACCGCTGGTGGCGCGCTCGCAACGCTGGACGAACGAGTC	1265
DB	149945	GCCGCACTGC-----GTCCGGAATTCGCCCAACAGACTGAGACGATGGAAGAGCGCATG	149998
QY	1266	CGCGAGCGCGCGAGCTGGCGCGGAGCGGATCGGTGCTGTTGTCGCGCGGCTCGCGG	1325
DB	149999	CGTTTGTGTCGCGCGACGTTCAAGCGGCTGATATGTTGCTGTTGTCGCGCGCTGGCGC	150058
QY	1326	AGCCTGGACATGTTCAAGAACTTCGAAGAACGCGGACGCTTTCGCCAAGACCGTAGAG	1385
DB	150059	AGCCTCGATCAGTTTAAATAATTTGAGCAACGGGCGGATGCTTTACCGCTGTGGCAAG	150118
QY	1386	GAGCTAGCGTGTATGC 1400	
DB	150119	GAGTTAGTGTATGC 150133	
RESULT 6			
AE016834			
LOCUS	300169 bp	DNA	linear BCT 21-MAR-2003
DEFINITION	Salmonella enterica subsp. enterica serovar Typhi Ty2, section 1 of 16 of the complete genome.		
ACCESSION	AE016834	AE014613	
VERSION	AE016834.1	GI:29136167	
KEYWORDS	Salmonella enterica subsp. enterica serovar Typhi Ty2		
SOURCE	Salmonella enterica subsp. enterica serovar Typhi Ty2		
ORGANISM	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Salmonella.		
REFERENCE	1 (bases 1 to 300169)		
AUTHORS	Deng, W., Liou, S. R., Plunkett III, G., Mayhew, G. F., Rose, D. J., Burland, V., Kodoyianni, V., Schwartz, D. C. and Blattner, F. R.		
TITLE	Comparative Genomics of Salmonella enterica Serovar Typhi Strains Ty2 and Cr18		
JOURNAL	J. Bacteriol. 185 (7), 2330-2337 (2003)		
MEDLINE	22531367		
PUBMED	12644504		
REFERENCE	2 (bases 1 to 300169)		
AUTHORS	Deng, W., Liou, S. R., Plunkett, G. III, Mayhew, G. F., Rose, D. J., Burland, V., Kodoyianni, V., Schwartz, D. C. and Blattner, F. R.		
TITLE	Direct Submission		
JOURNAL	Submitted (25-SEP-2002) Laboratory of Genetics, University of Wisconsin - Madison, 445 Henry Mall, Madison, WI 53706, USA		
FEATURES	Location/Qualifiers		
source	1. .300169		
	/organism="Salmonella enterica subsp. enterica serovar Typhi Ty2"		
	/mol_type="genomic DNA"		
	/strain="Ty2"		
	/serovar="Typhi"		
	/sub_species="enterica"		
	/db_xref="taxon:209261"		
	190. .255		
gene			

```
/gene="thrL"
/locus_tag="t0001"
190..255
CDS
/gene="thrL"
/locus_tag="t0001"
/notes="corresponds to STY0001 from Accession AL513382:
Salmonella typhi Ctl8"
/codon_start=1
/transl_table=11
/product="thr operon leader peptide"
/protein_id="AA067735.1"
/db_xref="GI:29136168"
/translation="MNRISTTTTITTTGNGAG"
337..2799
gene
/gene="thrA"
/locus_tag="t0002"
337..2799
CDS
/gene="thrA"
/locus_tag="t0002"
/notes="corresponds to STY0002 from Accession AL513382:
Salmonella typhi Ctl8; homoserine dehydrogenase I"
/codon_start=1
/transl_table=11
/product="aspartokinase I"
/protein_id="AA067736.1"
/db_xref="GI:29136169"
/translation="MRVLKFGGTSVANAERFLRVADILESNSROGOVATVLSAPAKIT
HLVAMIEKTIGGDALPNISDAERIFSDLLAGLASAOPGFLPLARLKMVVEQFAQIK
HVLHGSLGQCPDSNAALICRGEKMSIAGLLEARGHRTVYIDPVEKLVAVGHY
LESTVDIASTRIRASQIPADHMLMAGFTAGNEKGEVLVLRNGSDISAAVLAACL
RADCEITDWDVCTDPRQVDPARLLKMSYQEAELSIFGAKYVLRHTTPTIAOF
QPCLIKNTGNQAPGLIGASDDNLPVKGISLNNMFMFVSGPGMKMGIMGAAR
VFAAMRAGISVLIITQSSSEYSISFCPQSCARARMAQDEFYLELKEGLEPLAV
TERLAISVVGDMRTLRGISAKFFAALARANINIVAIAQSSERSISVVVNNDDATT
GVNTHQMLNDQVIEVGVIGVGVGALLEQLKQQTWLNKXHDIDVCGVANSKA
LLTNVHGLNDWQALANAPFNGLRILRLVKEHLNLPYIVCTSSQAVADQYAD
FLREGHVTPNNKANTSSMDYTHQRFAAQSRKFLYDNTNVAGLPLVIENQLNLL
AGDELQKFGISLGSLSFIFGLKEGMSLQATALAREMGYTEPDPDDLSDGMVARK
LITLARETGRELSDIVIEPVLDEFDASGDVTFPMHLQDLDFAAARVAKARDEG
NDVYAGTEEDGVCVKIAEVDGNDPLFKVKNGENALAFYSHYQPLPLVLRGYGAG
NVTYAGVADLLRLTSWLKV"
2801..3730
gene
/gene="thrB"
/locus_tag="t0003"
2801..3730
CDS
/gene="thrB"
/locus_tag="t0003"
/notes="corresponds to STY0003 from Accession AL513382:
Salmonella typhi Ctl8"
/codon_start=1
/transl_table=11
/product="homoserine kinase"
/protein_id="AA067737.1"
/db_xref="GI:29136170"
/translation="WKVYAPASSANMSVGFVLGAATVPDGTLLGDVVSVAEADHF
RLNHLGRFADKLPPRENIVYQWERFCQALGKTIPTVAMTEKNNPISGSLGSACS
VVAALVAMHECGKPLNDTRLALMGELEGRISGSHYDNVAPCFGLGMLMEENGI
ISQVPGFEWLVLPYGPVKVSTAEARAILPAQYRRQOCIAHGRHLAGFIHACYSRQ
POLAAAMKDVIAEPYRARRLLPGFSQARQAVSEIGALSGISGSGPTLALCDKPETA
QRVADWLKHYLQNSGFEVHCRLDTAGARVVG"
3734..5020
gene
/gene="thrC"
/locus_tag="t0004"
3734..5020
CDS
/gene="thrC"
/locus_tag="t0004"
/notes="corresponds to STY0004 from Accession AL513382:
Salmonella typhi Ctl8"
/codon_start=1
/transl_table=11
/product="threonine synthase"
/protein_id="AA067738.1"
/db_xref="GI:29136171"
/translation="MKLYNLKDHNEQVSAQAQVTOGLGKQQGLFFPHDLPEFSLTEID
EMLNQDFVSRSAKILSAFIDETPQOILREVRVAAFAFPAVAQVSDVGCLELFGHP
TLAFKDFGGRFMAQMLTHISGDKPVTILTSGDTGAANAHAAYGLENVVILYPRG
KISPLQKLFCTLGGNIEVADIDGDAQALVKQAFDDEELAKTALGLNSANSINISR
LIAQICVFEAAQALPQAGARNQLVISVPSGNFGDLTAGLLAKSLGLSPVRFYAAATNIN
DVPYRFDHGKWAQKATQATLSNAMDSQPNMNPVDELFRRKIWRILTEIGYAAVDOT
TTQOTMRELKAGYIIEPHAAVAYRALRQDLPNGEYGLFLGTAHPAKFKESVESILGE
TIALPEALAEADLPLLSHLHPADFAALRKLMMTRQ"
complement(5114..5887)
-gene
/gene="yaaA"
/locus_tag="t0005"
complement(5114..5887)
CDS
/gene="yaaA"
/locus_tag="t0005"
/notes="corresponds to STY0005 from Accession AL513382:
Salmonella typhi Ctl8"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AA067739.1"
/db_xref="GI:29136172"
/translation="MLILISPAKTYLQSPLATRYTQPELDDHSHOOLIQARQLSAP
QISRLMGISDKLADLNATRFHDMQPHFTPDNAQALAFKGDVITGLQAEITFNDADF
FAQQHLRMLSGLYLRLDLMQPYRLEMGIRLENPRKDLQYQFGDIITDKLINEALE
AQGDVYVNLASEEYFKSVKPKKDLAELIKPVFLDEKNGKFVYSYAKKARGLMRSF
IENRLTKPEQLTAFDREGYFDEETSTQDELVFKRYEQ"
complement(5966..7396)
gene
/gene="yaaJ"
/locus_tag="t0006"
complement(5966..7396)
CDS
/gene="yaaJ"
/locus_tag="t0006"
/notes="corresponds to STY0006 from Accession AL513382:
Salmonella typhi Ctl8"
/codon_start=1
/transl_table=11
/product="putative amino-acid transport protein"
/protein_id="AA067740.1"
/db_xref="GI:29136173"
/translation="MPEFFSFTNEILWGSVMYLLLAGCWFRTWRTGTFQFVYIROFS
RSIKLSLSPQGLTSTQALCTSLAARGSNLGAVALTAAGPGCAVFMWVSALIG
MATSFAECSLAQLYKERDPTGGPGPANTMARGLNRMWGVFALLFLVAGLIENS
VQANVSRALHFAFNIPPLISGLIALAFCALLIIRIGKVARLMQWLPLIALLLWVAG
SVFICLWHIEQMPGVIAISVKSFAFQMOEAAAGAGYTLTQAITSGFORGMFSNEAGM
STPNAATAATSYPPHPVAQGIQVMSIGVSDTIIICIVASAMILLAGNHASHSHTGEG
LIQHAWSLTGEWGSFVALIVLFAFSIVANYIYAENNLFLRCHNKAITWLLRLA
LTGMVITAGTLLISFPLIMQADMTMACMAITNLTAILLSPVYVTLAGDYLQRKLGVR
PQDPRFRFDIEPQLADPTWDAASRD"
7665..8618
gene
/gene="talB"
/locus_tag="t0007"
7665..8618
CDS
/gene="talB"
/locus_tag="t0007"
/notes="corresponds to STY0007 from Accession AL513382:
Salmonella typhi Ctl8"
/codon_start=1
/transl_table=11
/product="transaldolase B"
/protein_id="AA067741.1"
/db_xref="GI:29136174"
/translation="MTDKLTSRQFTTVVADTGDIAAMKLYQPDATTNPSSLTNAAQ
IPEYRKLIDDVAWAKQSSDRAQQVVDATDKLAVNIGLEIKLVPGRISTEVDARLS
YDTEASIAKAKRIKLYNDGRILKLASTWQIRAAEQLEKGINCLTLLFS
FAQARACAGVYVLIISPFVGRILDWYKANTDKDYAPEDPGVGVSVTEIYEVYKHGY
ETVWAGASFRNVGEILELAGCDRLTIAPALLKELAESEGAIERKLSFSSEVKARPERI
TBAEFUWHQHPMAVDKLDGIRKFAVDQEKLEKMGIDL"
8729..9319
gene
/gene="mog"
/locus_tag="t0008"
8729..9319
CDS
/gene="mog"
/locus_tag="t0008"
```

```
/note="corresponds to STY0008 from Accession AL513382:
Salmonella typhi CT18"
/codon_start=1
/transl_table=11
/product="molybdopter in biosynthesis Mog protein"
/protein_id="AA067742.1"
/db_xref="GI:29136175"
/transl_table="MNTLRIGLVSISDRSSGVYQDKIPALAEWLASALTTTPELOT
RLIPDEALIEQTLCELVDEMSHLVLTGTGTPARDVTPDATALADREMPFGBO
MROISLRFVPTAILSRQGVIRKQALILNLPQPKSIKETLEGKADDSVSPGIFA
SVYCIQLLDGPVVEPAEVAAPKPSARRENSS"
complement(9376..9942)
/gene="yaaH"
/locus_tag="t0009"
complement(9376..9942)
/gene="yaaH"
/locus_tag="t0009"
/note="corresponds to STY0009 from Accession AL513382:
Salmonella typhi CT18"
/codon_start=1

Query Match      23.28; Score 335.8; DB 1; Length 300169;
Best Local Similarity 56.28; Pred. No. 1.5e-31;
Matches 750; Conservative 0; Mismatches 552; Indels 33; Gaps 5;

Qy 69 GACCACCTTCGGCATCGTTGTCGGCTCGGCAAGAGCGGCATGTCCCTGGTGGCTACCTG 128
Db 148821 GACAAAACGTCGTCATTATCGGCTCGGCTTAACCGGACTCTCTTCGGTGGACTTTT 148880

Qy 129 GCGCGCGCGGCTTTCCTTCGCGCTGGTGCATACCCGAGAGAACCCGCGGAGCTGGCC 188
Db 148881 CTCGCCCGCGGCGTACGCGCGGCTGTATGATGATCTACGCGTACGCGCGGCTTCGAT 148940

Qy 189 ACCCTGCGTGCACGATATCCGAGTGAAGTGGTGGCGGAACTCGACGCGGAGTTC 248
Db 148941 AAGTCGCG-----CAAGAGTTGAGCGTCACGTTGGCGGCTTGAACGAGGATGG 148991

Qy 249 CTCCTGCTCCGCGCGCACTATGTCAGCCCGCGCTTGTGCTGCGCACCCCTCGCGCTG 308
Db 148992 CTCCTAGCGCGGATTTAATCGTCGCGCAGCCCTGTTATGCTCTGCGGCATCCCTCTT 149051

Qy 309 GTACAGCGCGCGGCGGAAAGGCTGCGCATCTCCGGTGCATGATCTTCGCGCGCGAG 368
Db 149052 AGCGCTCCGCGCAGCGCTGGAGTGAATTTGTCGGCGATATCGAACTGTTTTCGCGGAA 149111

Qy 369 GCGAAGCGCGCGATCGTCGCGCATCACCGGTTCCAAACGGAAGACACCGTGACACCCCTG 428
Db 149112 GCGCAAGCGCGGATGTGGGCATCACCGGCTCGAACGCGGAACACCGTGACACCCCTA 149171

Qy 429 GTGGCGGAAATGGCGGTGGCGGAGACAGCGTGTGCGCGTGGCGGCAACCTCGGCACC 488
Db 149172 GTGGCGGAGATGGCGGAGCGGCGGCGTCAATGTCGCGTGGCGGCAATATCGGTCCT 149231

Qy 489 CCGCGCTCGACCTGCTGGCGGAGCAGCATGAGCTGTACGTGTGGAGCTGTGCGAGTTC 548
Db 149232 CCGCGCTGTAGTGTGCTGGATGCGGATTCGCGGAATTTGACGTGTGGAATATTCAGTTTC 149291

Qy 549 CAGCTGAAACCTCGCATCGCTCAACCGCGAGTGGCGAGCGGCTGTAAGCTCAGCGAA 608
Db 149292 CAATTGGAACTACTCAAGTTTCAAGCGCGCGCGGCAAGCTGTCAAGCTCAGCTGAA 149351

Qy 609 GACCATATGATCGCTA---CGACGGCATGGCTGACTACCACTTGGCCCAAGCACCGGAT 665
Db 149352 GATCATATGACCGGATTCGCTTTGGTTTGAACAGTACCGGGGCGGAAACTCGCGGTC 149411

Qy 666 TTCGCGGTCGCCCGCAGTGTGCTGAATCGCGCCGATGCCCTGACCCGACCGCTGATC 725
Db 149412 TACGAAGCGGAAAGTGTGCGTAGTAGTGAATGCGATGACGCGTGTGACGATGCGGATCG 149471

Qy 726 GCCGATACCGCTGCTGCTGCTCGGCTGAACAAAGCGGAGCTTCAAGGCTTTCGCG 785
Db 149472 GGGCCGATGAGCGCTCGCTGAGCTTTGGCGTCAATATGGGTGATATCA-----C 149522
```

```
Qy 786 CTGATCGAGGAAGACGCCACAGAGTGGCTGGCGTTCCAGTTCGACAAAGCTCTCGCGGTT 845
Db 149523 CTTAATCGTCAGCAGCGAAACCTGCTACGAGTCAAAGTGAAGAGTCTGAACGTG 149582

Qy 846 GCGCAACTGAAGATCCGTGGCGGCCACAACTATTTCACACGGCTCGCGCGCTGCGCGTG 905
Db 149583 AAAGAGATGAAGCTTTCGGTTCAGCATTAACATATACGTTAGCGCGCTGCGCGTG 149642

Qy 906 GCGCATCGGTGGCGCTTCCGCTTCAGCGCCATGTCCGCGCGCTGAAGCGCTTTCCGCGC 965
Db 149643 GCGGATCGCGTAGTGTCTCCGCGCGCAGCAGTGTGAAGCGCTTGACGACATTTACCGT 149702

Qy 966 CTGCTCATCGTCCAGTGGGTACGCGAGCGGCGGCTGAGCTACTACGACGATTC 1025
Db 149703 CTGCGCATCGCTTCCAGCTGGCGTTGGAGCATTAACGCGCTTCTTGGATCAACGACTCG 149762

Qy 1026 AAGCCACCAACGTCGGCGCGCCCTGGCGCGGATCGAGGGCTGGGTGCGGACATCGAC 1085
Db 149763 AAACCGACCAATGTCCGCGAGTACCGAAGCGCGCTAAACGGTTTTCATG-----TGGAC 149816

Qy 1086 GGAAGCTGTGTGCTCGCGCGGAGACGCGCAAGGGCGCGGATTTCCATGACTCGCGC 1145
Db 149817 GGTAGCTGATCTGCTGCTGGCGGCGAGCGTAGTTCGCGAGACTTTTCTCCGCTGCGC 149876

Qy 1146 GAGCGGTGCGCGCTTCTGCGCGGCGGTGTACTGCTTGGCGCGTACGCGCGGCTGATT 1205
Db 149877 CGCTATCTGACCGCGATCGTATCGGCTGTATTGCTTTGGCGCGATGCGCGCAGCTT 149936

Qy 1206 GCCCAGCACTGGCAACGCGGTACCGCTGTGTGCGTGCAGCGTGGAGCGAGCAGTGC 1265
Db 149937 GCCCAGCTGC-----GTCCGGAATCGCCCAACAGACTGAGACGATGAAGAGCGGATG 149990

Qy 1266 CCGCAGCGCGCGAGCTGGCGCGGAGCGCATCGGTGCTGTGTCGCGCGCTGCGCG 1325
Db 149991 CGTTTGTGGCGCGCAGCTTACCGCGGTGATATGTCGCTGTGTCGCGCGCTGCGCC 150050

Qy 1326 AGCTGACATGTTTCAAGAACTTCAAGAACCGGACGCGCTTGTTCGCCAAAGCGTAGAG 1385
Db 150051 AGCTCGATCATGTTTAAATAATTTGAGCAACGGCGGATGCTTTACCCGCTGCGGAG 150110

Qy 1386 GAGTACGGTGATGC 1400
Db 150111 GAGTTAGTTGATGC 150125

RESULT 7
AE008699 Salmonella typhimurium LT2, section 7 of 220 of the complete genome. BCT 23-APR-2003
DEFINITION AE008699 AE006468
ACCESSION AE008699.1 GI:16418608
VERSION AE008699.1
KEYWORDS Salmonella typhimurium LT2
SOURCE Salmonella typhimurium LT2
ORGANISM Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
REFERENCE 1 (bases 1 to 22348)
AUTHORS McClelland, M., Sanderson, K.E., Spieth, J., Clifton, S.W.,
Latreille, P., Courtney, L., Porwollik, S., Ali, J., Dante, M., Du, F.,
Hou, S., Layman, D., Leonard, S., Nguyen, C., Scott, K., Holmes, A.,
Grewal, N., Mulvaney, E., Ryan, E., Sun, H., Florea, L., Miller, W.,
Stoneking, T., Nhan, M., Waterston, R. and Wilson, R.K.
Complete genome sequence of Salmonella enterica serovar Typhimurium
LT2
TITLE Nature 413 (6858), 852-856 (2001)
JOURNAL Nature 413 (6858), 852-856 (2001)
MEDLINE 21534948
PUBMED 11677609
REFERENCE 2 (bases 1 to 22348)
AUTHORS The Salmonella typhimurium Genome Sequencing Project
CONSRPM Direct Submission
TITLE Submitted (29-MAR-2001) Genome Sequencing Center, Department of
JOURNAL
```

Genetics, Washington University School of Medicine, 4444 Forest Park Boulevard, St. Louis, MO 63108, USA  
 COMMENT Supported by NIH grant 5U 01 AI43283

## COMMENT

Coding sequences below are predicted from manually evaluated computer analysis, using similarity information and the programs: GLIMMER; <http://www.tigr.org/softlab/glimmer/glimmer.html> and Genemark; <http://opal.biology.gatech.edu/Genemark/>

EC numbers were kindly provided by Junko Yabuzaki and the Kyoto Encyclopedia of Genes and Genomes; <http://www.genome.ad.jp/kegg/>, and Pedro Romero and Peter Karp at Ecocyc; <http://ecocyc.PangeaSystems.com/ecocyc/>

The analyses of ribosome binding sites and promoter binding sites were kindly provided by Heladia Salgado, Julio Collado-Vides and ReguonDB; [http://kinich.cifn.unam.mx:8850/db/regulondb\\_intro.frameset](http://kinich.cifn.unam.mx:8850/db/regulondb_intro.frameset)

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistries or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one m13 subclone.

## FEATURES

source  
 1. .22348  
 /organism="Salmonella typhimurium LT2"  
 /mol\_type="genomic DNA"  
 /strain="LT2; SGSC 1412; ATCC 700720"  
 /db\_xref="ATCC:700720"  
 /db\_xref="taxon:99287"  
 /note="LT2"  
 /gene="leuB"  
 /note="synonym: STM0110"  
 /complement(298. .903)  
 /EC\_number="4.2.1.33"  
 /note="3-isopropylmalate dehydratase small subunit."  
 (SW:LEU2\_SALTY)  
 /codon\_start=1  
 /transl\_table=11  
 /product="3-isopropylmalate isomerase (dehydratase), subunit with LeuB"  
 /protein\_id="AAL19074.1"  
 /db\_xref="GI:16418609"  
 /translation="NAEKFTQHTGLVPLDAANVDTDAIIPQFLQKVTTCFGAHLF  
 NDWRLEDEKQPNPEFVYNFPYQASILLARENFGCGSSREHAFWALTDYGFKVI  
 AFSFADFTGNSFNNQLLPVTUSDQVDFALVKANPGIKFEVDFLEAQVRAKDKTY  
 SFKIDDFRRHMLGLDSIGLTLQHEDAIAAYENQPAFMR"  
 /complement(910. .915)  
 /gene="leuB"  
 /note="putative RBS for leuB; RegulonDB:STMSIH000490"  
 /complement(914. .2327)  
 /gene="leuB"  
 /note="synonym: STM0111"  
 /complement(914. .2314)  
 /gene="leuB"  
 /EC\_number="4.2.1.33"  
 /note="3-isopropylmalate dehydratase large subunit."  
 (SW:LEU2\_SALTY)  
 /codon\_start=1  
 /transl\_table=11  
 /product="3-isopropylmalate isomerase (dehydratase), subunit with LeuB"  
 /protein\_id="AAL19075.1"  
 /db\_xref="GI:16418610"  
 /translation="NAKLYEKLFDHVVFEAPNETPLLYIDRHLVHEVTSPOAFDGL  
 RAHRPVPQKGTATMBDHNVTQTKDINASCAMIOEELKKNCEFGELYDLNH  
 PYQGLVHVMGPQGVTLPGMTIVCGDSHTATHGAFALAGIGTSEVHYLATQTLKQ  
 GRATKMEVYGNAAPIATKDVIATIGIKTGSAGGTGHVVEFCGDAIRALSMGEGMT  
 LCNMAIEMGAKAGLVAPDETTFNYYVKGRLHAPKGRDFDEAVEYKNTLTKDDGADFTY

VALRAEEIAPQVTWGTNPQGVISVTDIIPDPASFSDPVERASAEEKALAYMQLQPGVPL  
 TQVADIKVFIGSCTNSRIEDRAAAEAKRKVAPGVQALVVPGGSPVKAQAEAGLD  
 KIFIEAGFEWRLLPGGSMCLAMNRLNPGERCASNNRNFEGRGGRGTRHLVSPAMA  
 AAAAVTGHFADIRSIK"  
 /complement(2317. .3422)  
 /gene="leuB"  
 /note="synonym: STM0112"  
 /complement(2317. .3408)  
 /gene="leuB"  
 /EC\_number="1.1.1.85"  
 /note="3-isopropylmalate dehydrogenase. (SW:LEU3\_SALTY)"  
 /codon\_start=1  
 /transl\_table=11  
 /product="3-isopropylmalate dehydrogenase"  
 /protein\_id="AAL19076.1"  
 /db\_xref="GI:16418611"  
 /translation="MSKNYHIAVLPGDGGICPEVMAQALKAVMDAVSRFRDMRTTSHYD  
 VGGIADINDHGHPLPRATVEGCOADILFGSGPKWENLPPSPERGALLPLRRHKF  
 KLFNSLRPAKLYOGLAEFCPLRADTAANGFDILCVRELTTGGIYFGPKRGESGQYKE  
 AFDTEYHREFEIERIARIAFSARKKRRVTIDKANVLQSSILNREIYVNDVAKTYPD  
 VELAHMYIDNATMQLIKDPSQDFDLSNLFGLDILSDCAMITGSMGLPSASLNQOG  
 FGLYEPGAGSADPIAGKNANPFAOILSLALLRLYSLDANDAATAEQAINRALEBVG  
 RTGDLARGAAVSTDEMGDIARYVAEGV"  
 /complement(2322. .2327)  
 /gene="leuB"  
 /note="putative RBS for leuB; RegulonDB:STMSIH000491"  
 /complement(3408. .4999)  
 /gene="leuA"  
 /note="synonym: STM0113"  
 /complement(3408. .4979)  
 /gene="leuA"  
 /EC\_number="4.1.3.12"  
 /note="similar to E. coli 2-isopropylmalate synthase  
 (AAC73185.1); Blastp hit to AAC73185.1 (523 aa), 92%  
 identity in aa 1 - 523"  
 /codon\_start=1  
 /transl\_table=11  
 /product="2-isopropylmalate synthase"  
 /protein\_id="AAL19077.1"  
 /db\_xref="GI:16418612"  
 /translation="MSQQVIFDITLRGEQALQASLSAKEKLOLALALERMGVDMVE  
 VGFVSSPGDFESVQTIARTIKNSRVCALRCVENDIDVAQALKAVADAFRLHTFAT  
 SPMLATKLRLTDEVIERYIVNKRNYTDDEFSCEADGRTPVDDIARVVERAAIN  
 AGATINIPDTVTGTPPEFAGIISGLERVNIDKALISVHTHDGLAVGNSLAAY  
 HAGARVEGAMNGIGERAGNCALEEVMAIKYKIDIMNVHTNHNHINWRTSQVSI  
 CNMIPANKAIVGSAFAGSHGQDGLKKNREYIEMTPESIGINQILNLTSSGR  
 AAVKRMEMGVKDYNDMDHLYDAFLKADKKGVDFYDLAALAFINKQOEPEPHFR  
 LDYFSVQSSSDIATASVKLAGEIKAEANGPVDVIAIQAIRITGYQDVELVKYD  
 LNAKGGKDALQGVDIVVNNHGRFHGVLATDIVESSAKAMVHVLNINWRAAEVEKE  
 LQKKAQNKNNKETV"  
 /complement(3417. .3422)  
 /gene="leuB"  
 /note="putative RBS for leuB; RegulonDB:STMSIH000492"  
 /complement(4994. .4999)  
 /gene="leuA"  
 /note="putative RBS for leuA; RegulonDB:STMSIH000493"  
 /complement(5067. .5298)  
 /gene="leuB"  
 /note="synonym: STM0114"  
 /complement(5067. .5153)  
 /gene="leuB"  
 /note="leu operon leader peptide. (SW:LPL\_SALTY)"  
 /codon\_start=1  
 /transl\_table=11  
 /product="leu operon leader peptide"  
 /protein\_id="AAL19078.1"  
 /db\_xref="GI:16418613"  
 /translation="MSHIVRETGLLLNAFIVRGRPVGGIQH"  
 /complement(5162. .5167)  
 /gene="leuB"  
 /note="putative RBS for leuB; RegulonDB:STMSIH000494"  
 /complement(5271. .5279)  
 /gene="leuB"

-10\_signal

-35_signal	/note="putative -10_signal for leuL; RegulonDB:STMLTH004472" complement(5290..5298) /gene="leuL"	
gene	/note="putative -35_signal for leuL; RegulonDB:STMLTH004472" 5798..6754 /gene="leuO" /note="synonym: STM0115" 5798..5803 /gene="leuO" /note="putative RBS for leuO; RegulonDB:STMSI000495" 5810..6754 /gene="leuO" /note="probable activator protein in leuabcd operon. (SW:LEUO_SALTY)" /codon_start=1 /trans_table=11 /product="putative LysR family transcriptional regulator" /protein_id="AAL19079.1" /db_xref="gi:16418614"	
RBS	TLGSPQAVSNVAVLKVMEDELFRYGRIGIOTARAFOLFSVRQALQVONELPG SGFEPTSERVFNLCVSPDLNLTISQIYNRKIAPIHIVFRASLNQTEHOLRQ ETERVISTEERPREFTSVPLFKDEMVLVASRKHPRISGPLLEDGVNQEHAVYSLDR YASFQSPWYDTPDKOSSVAYQGMALISLVNVQSOTHLVAIPRWLAEEFAESLDLQIL PLPLKLSRTCYLSWHEAAGRDKGHWMDLLVSVCKR" 6941..6953	
protein_bind	Query Match 22.9%; Score 332.6; DB 1; Length 22348; Best Local Similarity 56.0%; Pred. No. 6.5e-31; Matches 748; Conservative 0; Mismatches 554; Indels 33; Gaps 5;	
Qy	69	GACCACTTCGCGATGTTGCGGCTCGGCAAGAGCGGCATGCTCCGTGGCTACCTG 128
Db	18656	GACAAAACGTCGTCATTTATTTGGCTTACCGGACTCTCTTGGCTGGACTTTT 18715
Qy	129	GCGCGCCGCGCTTTCGCGTTCGATCCCGAGAGAACCCCGCGAGCTGGCC 188
Db	18716	CTCGCCGCGGCTGACCGCGGCTGATGATCTCGCTGACGCGCGGCTGTGGAT 18775
Qy	189	ACCTGCGTCCAGTATCCGAGGTGAAGTGGTTCGCGGCACTGACGCGGATTC 248
Db	18776	AAGCTGCGG-----CAAGAGGTTGAGCGTCACGTTGGCGGCTGAACGACGAGTGG 18826
Qy	249	CTCTGCTCCGCGCGCAACTATGTACGCCCCGCTTGTGCTGCGGCAACCCCTCGCGTG 308
Db	18827	CTCTAGCGCGGATTTAATGTCGCGACGCTGTGTTGCTCTGCGGCATCCCTCTCTT 18886
Qy	309	GTACAGCGCGCGGAAAGCGTGGCGATCTCCGCTGACATCGATCTCTTCGCGCGGAG 368
Db	18887	AGCGCTGCGCGCAGCGCTGAGTGGAAATTCGCGGATATCGAACTGTTTTCGCGCAA 18946
Qy	369	GCGAAGCGCGCGATGCTGCGCATCACCGGTTCCAGCGAGAGACCGGTACCCACCTG 428
Db	18947	GCGCAGCGCGGATGTTGGCGCTCACCGCTCGAAGCGCAAGACCGGTGACCCACTTA 19006
Qy	429	GTGGCGAAATGGGCGTGGCGCGGACAAAGGCTGTGCGCGTGGCGGCAACCTCGGCACC 488
Db	19007	GTGGCGAGATGGCGAAGCGCGGCGTCAATGTGCGCGTGGCGGCAATATCGGTCTT 19066
Qy	489	CGGCGCTGACGTGCTGGCGGACGACATCTCCGCTGACATCGATCTGTGGAGCTGTGAGCTTC 548
Db	19067	CCGCGCTGATGCTGCTGGATGCGGATCGCGAATGTTACGTGCTGGAATATTCCAGTTTC 19126
Qy	549	CAGCTGGAAACCTGGATCGCTCAACCGCGAGTGGCGGACCGTCTCAAGCTACGCCAA 608
Db	19127	CAACTGGAAATACCTCAAGTTTTCGACGCGCGCGGCAACGGTGTCAAGCTCACTGAA 19186
Qy	609	GACCATATGGATCGCTA---CGACGGCATGCTGACTACCACTTGGCCCAAGCACCGGATC 665
Db	19187	GATCATATGACCGGTATCCGTTGTTTGGCAACGATACCGGGCGGCGGAACTCGCGCTC 19246
Qy	666	TTCCGCGTGGCCGCCAGGTGCTGTGAATCGCGCGGATGCCCTGACCGACCGCTGATC 725
Db	19247	TACGAGAAGCGCAAGTCTGCGTAGTGAATCCGATGACGCGTTCACGATCCGCTACGC 19306
Qy	726	GCGGATACCGTGGCTGCTGTGCTGCGCTGAAACAAGCGGACTTCAAGGCTTTCGCGC 785
Db	19307	GGGCGGATGAACGCTGCTGTAGCTTTGGCGCTCAATATATGGCGATTATCA-----C 19357
Qy	786	CTGATCGAGGAGAGCGGCAGAGTGGCTGGGTTCCAGTTCGACAAGCTGCTGCGGTT 845
Db	19358	CTTATTCGTCAGAGGGGCAACCTTGGCTGGAGTCAAAAGGGAAGAGTGTGAACGTG 19417
Qy	846	GCGCAACTGAAGATCCGTGGCGCCCAACAACCTATTCCAAAGCGGCTCGCGCGCTGCGCTG 905
Db	19418	AAAGAGATGAGCTTTCCGCTGACGATTAACATATACCAATGCTTGAAGCGCGCTGCGCTG 19477
Qy	906	GCGCATGCGGCTGCGCTGCGTTCGACGCCATGCTCGCGCGGCTGAAGCGCTTTTCGCGC 965
Db	19478	GCGGATGCGGTAGTCTGCGCGCGCGCAGCATTTTGAAGGCGTTCACGACATTTTACCCTG 19537
Qy	966	CTGGCTCATCTGCTGCGCTGCTGCGGCGGCTAGCGGCGGCGGCTGAGCTACTACGACATTC 1025
Db	19538	CTGGCGATCGCTTCCACTGCGCTTGGAGCATACGCGGTACGTTGGATCAACGACTCG 19597
Qy	1026	AAGGCCAACACGTCGGCGCCCGCTCGCGCGGATCGAGGGCTGGGTGCCGACATCGAC 1085
Db	19598	AAAGCGACCAACGTCGGCGAGTACCGAAGCGCGCTTAAACGGTTTTCATG-----TGGAC 19651
Qy	1086	GCGAAGCTGCTGCTGCTCGCGGGGAGACGCGCAAGCGCGGATTTCCATGACCTGCGC 1145
Db	19652	GGTACGCTACATCTGCTGCTGCGCGGCGACGAGTAAGTCGCGAGACTTTTCTCCGCTGACA 19711
Qy	1146	GAGCGGTCGCGCGCTTCTGCGGCGGTGCTGCTTGGCGGTGACGCGGCGCTGATT 1205
Db	19712	CGCTATCTGACGCGGCTGATTCGCGCTGTATTCCTTTGGCGGCGATGGCGCGAGCTT 19771
Qy	1206	GCCAGGCACTGGGCAACGCGTACCGTGTGCGCTGCGCAACCTCGAAGCAACGACTG 1265
Db	19772	GCGGCACTGC-----GTCCGGAATTCGCCAACAGACTGAGACGATGAAGAGCGCATG 19825
Qy	1266	GCGGAGCGCGCGAGCTGGCGGAAAGCGGATGCGGTGCTGTGCTGCGCGCGCTGCGCG 1325
Db	19826	CGTTGCTGGCGCGCGGCTTCAGCGGCTGATATGTTGCTGCTGCTGCGCGCTTTCGCT 19885
Qy	1326	AGCTTGGACATGTTCAAGAACTTCGAAGAAGCGGACGCGCTGTTCCGCAAGACCGTAGAG 1385
Db	19886	AGCTTCGATCATGTTTAAATAATTTTGAGCAACGGGCGATGCTTTACCGCTGCGCGAAG 19945
Qy	1386	GAGCTAGCGTGAATC 1400
Db	19946	GAGTTAGTTGATGC 19960
RESULT 8	AE015046 11441 bp DNA linear BCT 18-OCT-2002	
AE015046	Shigella flexneri 2a str. 301 section 9 of 412 of the complete genome.	
LOCUS	AE015046 AE005674	
DEFINITION	Shigella flexneri 2a str. 301	
ACCESSION	AE015046.1 GI:24050283	
VERSION	Shigella flexneri 2a str. 301	
KEYWORDS	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;	
SOURCE	Enterobacteriaceae; Shigella.	
ORGANISM	1 (bases 1 to 11441)	
REFERENCE	Jin, Q., Yuan, Z.H., Xu, J.G., Wang, Y., Shen, Y., Lu, W.C., Wang, J.H.,	
AUTHORS	Liu, H., Yang, J., Yang, F., Qu, D., Zhang, X.B., Zhang, J.Y., Yang, G.W.,	
	Wu, H.T., Dong, J., Sun, L.L., Xue, Y., Zhao, A.L., Gao, Y.S., Zhu, J.P.,	
	Kan, B., Chen, S.X., Yao, Z.J., He, B.K., Chen, R.S., Ma, D.L.,	
	Qiang, B.Q., Wen, Y.M., Hou, Y.D., and Yu, J.	
TITLE	Genome sequence of Shigella flexneri 2a: Insights into pathogenicity through comparison with genomes of Escherichia coli	

K12 and O157  
Nucleic Acids Res. 30 (20), 4432-4441 (2002)  
12384590  
2 (bases 1 to 11441)  
Jin, O., Shen, Y., Wang, J. H., Liu, H., Yang, J., Yang, F., Zhang, X. B.,  
Zhang, J. Y., Yang, G. W., Wu, H. T., Dong, J., Sun, L. L., Xue, Y.,  
Zhao, A. L., Gao, Y. S., Zhu, J. P., Chen, S. X., Yao, Z. J., Wang, Y.,  
Lu, W. C., Qiang, B. Q., Wen, Y. M., and Hou, Y. D.  
Direct Submission  
Submitted (08-MAY-2001) Microbial Genome Center of Chinese Ministry  
of Public Health, 100 YingXin Jie, XuanWu Qu, Beijing 100052, P.R.  
China

FEATURES  
source  
1. 11441  
/organism="Shigella flexneri 2a str. 301"  
/mol\_type="genomic DNA"  
/strain="301"  
/serotype="2a"  
/db\_xref="taxon:198214"  
292..786  
/gene="yabB"  
/locus\_tag="SF0078"  
292..786  
/gene="yabB"  
/locus\_tag="SF0078"  
/notes="Residues 13 to 164 of 164 are 99 pct identical to  
residues 1 to 152 of a 152 aa protein from Escherichia  
coli O157:H7 EDL933 ref: NP\_285777.1"  
/codon\_start=1  
/transl\_table=11  
/product="orf, conserved hypothetical protein"  
/protein\_id="AA041743.1"  
/db\_xref="GI:24050284"  
translation="MGKVGIRGEAGMFGATVNLDSKRLSPVTRYEOLLENAAG  
QMYCTDIDHPCLLLYLPPEWIEIQKLSLLSMNRYVRRVQRLILGHASECOMDAG  
RLIAPVQLQHAGLTKEVLMVGQFNKFLWDTWHQQVEDIAEQLATGDLSERLQ  
DLSL"  
788..1729  
/gene="yabC"  
/locus\_tag="SF0079"  
788..1729  
/gene="yabC"  
/locus\_tag="SF0079"  
/notes="Residues 1 to 313 of 313 are 99 pct identical to  
residues 34 to 346 of a 346 aa protein from Escherichia  
coli pir: QBCFt"  
/codon\_start=1  
/transl\_table=11  
/product="yabC protein"  
/protein\_id="AA041744.1"  
/db\_xref="GI:24050285"  
translation="MNEYKHTTVLLDEAVNGLNIRPDGIYIDGTFGRGHSRLILSQ  
LGEGRLAIDRDPQAIYAKVTTIDDPFSIIHGPFSAIGEYVAERDLICKIDGILLDL  
GVSPOLDAERGSEFMRDGLMDMDPTRGQSAEWLQTAEDIANVLYKYEHEHF  
AKRIARIVNRRQEPMTKELAEVVAATPVKDKFHPATPTFOAVRIWNSELEE  
IEQALSSUNVAPGGKRLSIIISPHSLDRIVKRFMRNSRGPPQVAGLPMTPEQLKKL  
GGQRLAKGLKMGEEVAENPARSVLRIARTNA"  
1726..2091  
/gene="ftsL"  
/locus\_tag="SF0080"  
1726..2091  
/gene="ftsL"  
/locus\_tag="SF0080"  
/notes="Residues 1 to 121 of 121 are 100 pct identical to  
residues 1 to 121 of a 121 aa protein from Escherichia  
coli O157:H7 EDL933 ref: NP\_285779.1"  
/codon\_start=1  
/transl\_table=11  
/product="cell division protein; ingrowth of wall at  
septum"  
/protein\_id="AA041745.1"  
/db\_xref="GI:24050286"  
translation="MISRYTEALSKVKGSMGSHERHALPGVIGDILLRFGKLPCLFI

gene  
/gene="ftsI"  
/locus\_tag="SF0081"  
2107..3873  
/gene="ftsI"  
/locus\_tag="SF0081"  
/notes="Residues 1 to 588 of 588 are 99 pct identical to  
residues 1 to 588 of a 588 aa protein from Escherichia  
coli O157:H7 EDL933 ref: NP\_285780.1"  
/codon\_start=1  
/transl\_table=11  
/product="septum formation; penicillin-binding protein 3;  
peptidoglycan synthetase"  
/protein\_id="AA041746.1"  
/db\_xref="GI:24050287"  
translation="MKAATQKPKQEEHANFISWRFALLCCILLALAFILGRVAM  
LQVISPDLVKEGDMRSLRVQVSTSGMITDRSGRLAVSPVPAKIMADPKREVHDAG  
GISVGDWALANALNIPDLQLSARINANPKGRFIYLAQVNSDMADYTKKLPGTH  
LREESRRYPSGGEVTAHLIGFTNVDSQGETGKSPDKWLTGQGERIVRKDRYGRVI  
EDISSTDSQAANHLASIDERQALYRELNNAVAKAESGSAVLVDNTGCVLAMA  
NSPSYNNLSGTPKEMRNRTIIDVFEGSTVYKPMVYMTALQRGVVRNSVLTPIPI  
RINGHEIKDARYSELTGTGVQKSSNVGSKLATMPSALSALVDITYRFGKATNLG  
LVYERGLYQKQWSDIERATFSFGYGLMTVPLQIARYATIGSYGIRPLSITVD  
PPVGRVFPESIVRTVWHMESVLPVGGGKAAIKGYRIAKITGTAKKVGPDGRYI  
NKVIYTAGVAPASQRFALVVINDPQAGKYVGAVSAPVFGAINGGVLRTWNIEPD  
ALTGDKNEVINQSGEGTGRS"  
3860..5347  
/gene="murE"  
/locus\_tag="SF0082"  
3860..5347  
/gene="murE"  
/locus\_tag="SF0082"  
/EC\_number="6.3.2.13"  
/notes="Residues 1 to 495 of 495 are 99 pct identical to  
residues 1 to 495 of a 495 aa protein from Escherichia  
coli pir: S40595"  
/codon\_start=1  
/transl\_table=11  
/product="UDP-N-acetylmuramoylalanyl-D-glutamate--2,  
6-diaminopimelate ligase"  
/protein\_id="AA041747.1"  
/db\_xref="GI:24050288"  
translation="MADRNLROLLAPWPDPAPSRLREMTLDSRVAAGDLFVAVVGH  
QADGRYIPQAIAGVAAIARAKDEATDGETIREMHGVPIVYLSQNERLSALAGRFY  
HEPSDNRLVGYTGNGKTTTQLLAQWSQLGETSAVMGTGNGLLGVIPENTTG  
SAVDQHELAGLVDDQATFCAMEVSHGLVQHRVAALKFPASVFTNLSRDHLDYHGD  
EHYEAAKWLLYSEHCGQAIINADDEVRRRLAKLPDAVAVSMEDHINPNCGRWLKA  
TEVNYHDSGATIRFSSGWDGETESHMGAFVSNLLALATLALGYPLADLLKTA  
RLQPVGRMEVTPAGKTVVYVYDHTPDALKAQARLHCAGKLCVFGCGGDRDK  
GKPLMGATAEEFADVAVVYDDNPRTEPEPRAIINDLAGMLDAGHAKVMEGRAEAVTC  
AWQAKENDVVLVAGKHEDYQIVGNQLRDISDRVTVARLLGGIA"  
5344..6702  
/gene="murF"  
/locus\_tag="SF0083"  
5344..6702  
/gene="murF"  
/locus\_tag="SF0083"  
/notes="Residues 1 to 452 of 452 are 99 pct identical to  
residues 1 to 452 of a 452 aa protein from Escherichia  
coli K12 ref: NP\_414628.1"  
/codon\_start=1  
/transl\_table=11  
/product="D-alanine:D-alanine-adding enzyme"  
/protein\_id="AA041748.1"  
/db\_xref="GI:24050289"  
translation="MISVTLSQLTDILNGLQAGADITLDAVTTDTTRKLPFGCLFVALK  
GRFPAHDFAQKAGAGGAGALLVSRLDIDLPLQIVKDRTRFLAFGELAAVVRQVPAIV  
VLTGSSGTSYKVGNTLAALSSQGNLTLYAGLNNDIGVPMTLRLTPEYDYAVIELG  
ANHGEIAWTVSLTRPEAALVNNLAAHLEGFSGLAGVAKAGEIFSGLPENGICIAMN  
ADNDWLNQSVISGRKVMRFSPNANSDFATNTHVTSHGMEFTLQPTSGVSVLLP  
LPGRRHNIALAAAAALSMVSGATLDAIKAGLANLKNKVPGRLPFIQLAENQLLLDSSYN



ANVSGMTAAVQVLAEMPGYRVLVVGDMAELGAESEACHVQVGEAAKAAGIDRVLSMGK  
 QSHAISTASGVGEHFADKLTALITRLKSLIAEQVITILVKGSRSAAMEEVVRLQENG  
 TC  
 6696..7778  
 /gene="mray"  
 /locus\_tag="SF0084"  
 6696..7778  
 /gene="mray"  
 /locus\_tag="SF0084"  
 /note="Residues 1 to 360 of 360 are 100 pct identical to  
 residues 1 to 360 of a 360 aa protein from Escherichia  
 coli K12 ref: NP\_414629.1"  
 /codon\_start=1  
 /transl\_table=11  
 /product="phospho-N-acetylmuramoyl-pentapeptide  
 transferase?"  
 /protein\_id="AA041749.1"  
 /db\_xref="GI:24050290"  
 /translation="MDVWLAELHVKYVSGFNVSFLYFRFAIVSLTLALFISLMGPRM  
 IAHQLKLSFGQVVRNDGPESEHFSKRGTPTMGIMILTAIVISVLLWAPSPNYYWCVL  
 VLVYGVIGVDDYRKVRKDTGLIARWKYFWSVIALGVAFALVAGDKDTPATOL  
 VVPFKDVPOLGFYILLAVFVIVGTGNVNLTDGLDGLAIMPVFAVAGSFALVANA  
 TGNNEFASLHPIYLRAGELVIVCTAIVGAGLGFLEWNTYPAQVFDVGSALGGA  
 LGIATVLRQELVINGGVFVETLSVILQVGSFKLRGQIFRMAPLHHHYELKQWP  
 EPRVIVRFRWILSLMLVLIQATLKV"  
 7781..9097  
 /gene="murD"  
 /locus\_tag="SF0085"  
 7781..9097  
 /gene="murD"  
 /locus\_tag="SF0085"  
 /note="Residues 1 to 438 of 438 are 99 pct identical to  
 residues 1 to 438 of a 438 aa protein from Escherichia  
 coli K12 ref: NP\_414630.1"  
 /codon\_start=1  
 /transl\_table=11

Query Match 20.1%; Score 291.2; DB 1; Length 11441;  
 Best Local Similarity 53.6%; Pred. NO. 6.2e-26;  
 Matches 736; Conservative 0; Mismatches 603; Indels 33; Gaps 5;

QY 82 TCGTTGTGGCGCTGGCAAGAGCGCATGTCCTTGGTGGCTACTGGCGCGCGCGCT 141  
 Db TCAATATGGCGCTGGCGCTACCGGGCTTCTCGTGGAGCTTTTCTCGCTCGCGGTG 7868  
 QY 142 TGCCTTTTCGCGTGTGATACCCGAGAAACCGCGGAGCTGGCCACCTCGCTGCC 201  
 Db TGACCGCGCGGTATGATACGCGTATGACACCGCTGGCTGGATAAATACCG--- 7925  
 QY 202 AGTATCCGAGTGAAGTGGTTCGCGGAACTCGACCGCGAGTTCCTCTGCTCGGCC 261  
 Db TTTT-----AAGCGGTAGAACGCGCACACGCGCAGTCTGAATGATGAATGGCTGATGGCGGAG 7979  
 QY 262 CGGAACCTATGTACGCGCGCGCTGTGCTGGCGACCCCTGCGCTGTACAGCGCGCG 321  
 Db ATCTGATTTGTCGCGATCCCGGTATGACACTGGCGCATTCATCTTAAGCGCTCGCGGTG 8039  
 QY 322 CGAAGAGCGGTGGCATCTCCGCTGACATCGATCTTTCGCGCGGAGGCGAGGCGCCGA 381  
 Db ATGCGGGAATCGAATCTTGGCGATATCGAGCTGTTTTCGCGGAGCAACAGCACCGA 8099  
 QY 382 TCGTGGCATCATCGGTTTCCAAACGGAAGACCGTGGACCACTGCTGGTGGCGCAATGG 441  
 Db TTGTGGCATTACCGGTTTCAACGCGCAAAAGACCGTCAACCACTAGTGGTGAATGG 8159  
 QY 442 CGGTGGCGCGGACAGCGTGTGCGGCGGCAACTCGGCACCCCGCGGTGCGAC 501  
 Db CGAAAGCGCGGGGTTTAACTGTGTGTGGTGGCAATATTCGCTGCTGCTGTGATGC 8219  
 QY 502 TCGTGGCGGACATCGAGTGTACGTGTTGGAGCTCTCGAGCTTCCAGCTGGAACT 561  
 Db TACTGGATGATGATGTGAACGTACGTGCTGGAACCTGCTGAGCTTCCAGCTGGAACT 8279

562 GCATCGCTCAACGCGGAGGTGGCGCGTGTGAACGTCAGCGAAGACCATATGGATC 621  
 Db CTCCAGCTTACAGCGGTAGCAGCGACCATTTGAACGTGACTGAAGATCATATGGATC 8339  
 QY 622 GCTAC---GACGGCATGGCTGACTTACACCTGGCCAAAGCAGCGGATCTTCCGCGGTGCC 678  
 Db GCTACCGGTTTGTATTACACAGTATCGTGCAGCAAACTCGCATTTTACAAAAACGGA 8399  
 QY 679 GCCAGGTGCTGTGTAATCGCGCGCATGCCCTGACCGACCGCGCTGATCGCGCATACCGTGC 738  
 Db AAGTTTGGGTGTTAATGCTGATGCTTAAACAAATGCCGATTCGCGGTGCGGATGAAC 8459  
 QY 739 CGTCTGGTCTGCTCGCGCTGAACAAAGCGCGACTTCAAGGCTTTTCGGCTGTATCGAGGAAG 798  
 Db GCTCGCTGAGCTTTGGCGTCAACATGGGTGACTATCA-----CCTGAATCATCAGC 8510  
 QY 799 ACGGCCAGAGTGGCTGGCTTCCAGTTCGACAGCTGCTGCCGTTGGCGCAATGAAGA 858  
 Db AGGCGCAAACTGGCTGGCTGCGGTAAAGCGGAGAAAGTCTGAATGTGAAGAGATGAAC 8570  
 QY 859 TCCGTGGCGGCCCAACTATTCCAAACGCGCTCGCGCTGGCGTGGCGCATCGCGTGC 918  
 Db TTTCCGGGCGACATACATACCAATGGCTGGTGGCTGGCGTGGCGTGGCGATGCCGAG 8630  
 QY 919 GCCTGCCCTTTCAGCCCATGCTCGCGCGCTGAAGGCGTTTTCGGCTGCTCATCGCT 978  
 Db GTTTACCGCGCGCGCAGCAGCTGAAAGCGTTTAAACACCTTCCACCGTCTGCCGATCGCT 8690  
 QY 979 GCCAGTGGGTAGCGGAGCGGCGTGGCTGACTACTAGCAGTTCAGGCGTCCAAAGCCACCAACG 1038  
 Db TTGAAGTGGTCTGGAGCATTAACGCTGGCTGGATTAACGATTAACAAAGCGACCAACG 8750  
 QY 1039 TCGCGCGCGCTCGCGCGCATCGAGGCGTGGTGGCGACATCGACGGAAGCTGGTGC 1098  
 Db TCGCGAGTACGGAAGCGCGCTGATGGCTG-----CAGTAGACGCGCGCTGCATT 8804  
 QY 1099 TGTTCGCGCGGAGAGCGCAAGGCGCGCTGCAAGCTGCAAGAGCTGCGGAGCGGTCGCGC 1158  
 Db TGTTCGCGCGCGCATGTTAAATCGCGGACTTTAGCCCACTTGTGCTGCTTACCTGAATG 8864  
 QY 1159 GCTTCTCGCGCGGTGTACTGCTTGGCGCTGACCGCGGCTGATTCGCCAGGCGACTGG 1218  
 Db GCGATAAGTACGTCTGTTTTCGCTGCTGACGCGCGCAGCTGGCGGCGCTAC--- 8921  
 QY 1219 GCAACGCGGTACCGCTGGTGGCGCTGCAAGCTGCAAGAGCTGCGGAGCGGCGCGCG 1278  
 Db ---GCCGGAAGTGGCAACAACACCAACTGGAACAGCTATGCGCTTGTGGCTC 8978  
 QY 1279 AGTGGCGCGGAGAGGCGATCGGTGCTGTGTCGCGCGCTGCGGAGCGCTGGACATGT 1338  
 Db CGCGTGTTCAGCGCGCGATATGTTCTGCTCTCCCGCGCTGTGCCAGCGCTTGCATCAGT 9038  
 QY 1339 TCAAGAACTTCAAGAACGCGGCGCGCTTTCGCGCAAGCGCTAGAGAGCTAGCGTGCAT 1398  
 Db TCAAGAACTTGAACACAGGCGCAATGATTTTCCCGCTTGGCGAAGAGTAGTGTGAT 9098  
 QY 1399 GCTGTGCTGTGTGGCGCGCTTCCCGCTGCGCGCTGTTAGCGCGCACGCGCATC 1450  
 Db GCGTTTATCTCTCCCTGCCCTGAAAATGCCGCGCTGCCAGATTCAATATC 9150

RESULT 9  
 AE016978  
 LOCUS Shigella flexneri 2a str. 290029 bp DNA linear BCT 22-APR-2003  
 DEFINITION Shigella flexneri 2a str. 2457T section 1 of 16 of the complete  
 genome.  
 ACCESSION AE016978  
 VERSION AE016978.1  
 GI:30039813  
 KEYWORDS Shigella flexneri 2a str. 2457T  
 SOURCE Shigella flexneri 2a str. 2457T  
 ORGANISM Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 Enterobacteriaceae; Shigella.

```

REFERENCE
AUTHORS
1 (bases 1 to 290029)
Wei,J., Goldberg,M.B., Burland,V., Venkatesan,M.M., Deng,W.,
Fournier,G., Mayhew,G.F., Plunkett,G. III, Rose,D.J., Darling,A.,
Mau,B., Perna,N.T., Payne,S.M., Runyen-Janecky,L.J., Zhou,S.,
Schwartz,D.C. and Blattner,F.R.
Complete Genome Sequence and Comparative Genomics of Shigella
flexneri Serotype 2a Strain 2457T
Infect. Immun. 71 (5), 2775-2786 (2003)
JOURNAL
PUBMED
12704152
REFERENCE
AUTHORS
2 (bases 1 to 290029)
Wei,J., Goldberg,M.B., Burland,V., Venkatesan,M.M., Deng,W.,
Fournier,G., Mayhew,G.F., Plunkett,G. III, Rose,D.J., Darling,A.,
Mau,B., Perna,N.T., Payne,S.M., Runyen-Janecky,L.J., Zhou,S.,
Schwartz,D.C. and Blattner,F.R.
Direct Submission
Submitted (13-JUN-2002) Genetics Laboratory, University of
Wisconsin - Madison, 445 Henry Mall, Madison, WI 53706, USA
FEATURES
source
1. .290029
/organism="Shigella flexneri 2a str. 2457T"
/mol_type="genomic DNA"
/strain="2457T"
/serotype="2a"
/db_xref="taxon:198215"
190. .255
/genes="thrL"
/locus_tag="S0001"
190. .255
/genes="thrL"
/locus_tag="S0001"
/locus_tag="S0001"
/function="leader; Amino acid biosynthesis: Threonine"
/notes="residues 1 to 21 of 21 are 100 pct identical to
residues 1 to 21 of 21 from Escherichia coli K-12 : B0001"
/codon_start=1
/transl_table=11
/product="thr operon leader peptide"
/protein_id="AAP15348.1"
/db_xref="GI:30039814"
/translation="MKRISTITTTITTTTNGAG"
336. .2798
/genes="thrA"
/locus_tag="S0002"
336. .2798
/genes="thrA"
/locus_tag="S0002"
/function="enzyme; Amino acid biosynthesis: Threonine"
/notes="residues 1 to 820 of 820 are 98.04 pct identical to
residues 1 to 820 of 820 from Escherichia coli K-12 :
B0002"
/codon_start=1
/transl_table=11
/product="aspartokinase I, homoserine dehydrogenase I"
/protein_id="AAP15549.1"
/db_xref="GI:30039815"
/translation="MRVLFQGSTVANAERFLRVADILLESNAQGVATVLSAPAKIT
NHLVMIENITSGDALPNISGAEIRFASLLGLAAQPGFLAQKTFVDQFEAQIK
HVLHGISLGGQCPDSINRAILGKEMSTIAMGVLKARGHNTVIDPVKLLAVGHY
LESTVDIALENKWEQAQAKPEFNLGRLIRLVKEVFLNPLVICTSSQAVADQYAD
RADCCFIWTDVGVYCDPQVDPARLLKMSYQEAELSYFGAKVLHPRTTTPAOF
QIPCLIKNTGNPQAGTLIGASRDEDELVPKGISLNNMAMFESVSGPMKGMVGAAR
VFAMRSARISVVLITQSSEYSISFCVPQSDCVRAERAMQEEFYLEKEGLEPLAV
TERLAISVVGMRTRLRGISAKFFAALRANINIVAIQSSERSISVVVNDQATT
GVYTHQMFNTDQVLEVEVIGVGGGALLQLKQOQSLKKNKHIDLRVCVANSKA
LITSVHGLNKNWEQAQAKPEFNLGRLIRLVKEVFLNPLVICTSSQAVADQYAD
FLREGPHVTPNKNKANTSMQYHOLRYAAEKSRKFLYDVTNAGLPIVENLQNLN
AGDELVKFSGLISLSYIFGKLDCGMSFSEATLTAREMGYTEPDPRDLSQMDVARK
LILATREGLRELADIEIEPLVPREFNAGDVAAFMANLSQLDDLFAARVAKARDEG
KVLRYVGNIDEGVCRKIAEVDNDPLFKVNGENALAFYSHIYQIPLPLVRGYGAG
NDVTAGVFADLLRTLTSWKLGV"
2800. .3732
/genes="thrB"
/locus_tag="S0003"
2800. .3732
/genes="thrB"
/locus_tag="S0003"
/locus_tag="S0003"
/function="enzyme; Amino acid biosynthesis: Threonine"
/notes="residues 1 to 310 of 310 are 95.48 pct identical to
residues 1 to 310 of 310 from Escherichia coli K-12 :
B0003"
/codon_start=1
/transl_table=11
/product="homoserine kinase"
/protein_id="AAP15550.1"
/db_xref="GI:30039816"
/translation="MVKVYAPASSANMSYGFVILGAAPVADGALIGLVVTEAAETFF
SNNLRADFKLPSEPRENVIVQWRFQELGKQIPVAMTLEKNMPICGSLSSACS
VVAALMAMNEHCGKPLNDRLIALMGELGRISGSIHYDNVAPCFGGMQLMEENDI
ISQOVPGFDEWLVLVAYPGIKYKSTAEARAILPAQYRQDQCIAGHRLAGFIHACYSRQ
PELAAKLMDVIAEPYRERLLPGFRQARQAVAEIGAVASGISGSGPTLALCDKPDPA
QRVADWLIGNYQNGEFGVHICLKDITAGARVLEN"
3733. .5019
/genes="thrC"
/locus_tag="S0004"
3733. .5019
/genes="thrC"
/locus_tag="S0004"
/function="enzyme; Amino acid biosynthesis: Threonine"
/notes="residues 1 to 428 of 428 are 95.56 pct identical to
residues 1 to 428 of 428 from Escherichia coli K-12 :
B0004"
/codon_start=1
/transl_table=11
/product="threonine synthase"
/protein_id="AAP15551.1"
/db_xref="GI:30039817"
/translation="MKLYNKDHNEQVSPAQVTOGLKNGQLFFPHDLPEFSLTEID
EMKLDIVFRSAKILSAFIDELPOEILREVRAPAFAPVAVNVEDVGCLELPHGP
TLAFDEGGFRMAQMLTHIADDPVTILTATSDTGAAVAHAFLGPNKVVILPRG
KISPLQEKFLTLGNIEVAIDGDQACALVKQAFDDEELKVALGLNSANISLR
LIAQICYFEVAQVQEARNLQVSVPSNGFDLTAGLLAKSFGLPVKRFIAATNVN
DTVPRFLHDGQSPKATQATLSNMDVQPNMNPVRELFRRKINQKELGFAAVDDE
TTQOTMRELKELGYTSEPHAAVAYRALQDLNPGEYGLFGTAHPAKFESVEAILGE
TLDLPKELAEADLPLLSHNPADFSAKRKLMMNHQ"
5242. .5529
/locus_tag="S0005"
5242. .5529
/locus_tag="S0005"
/notes="residues 1 to 65 of 95 are 60.00 pct identical to
residues 4 to 68 of 98 from Escherichia coli K-12 : B0005"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAP15552.1"
/db_xref="GI:30039818"
/translation="MQSIVLALSVLVAPMAAAAEITLVPSVKLQIGDRDNRGVYWD
GGHWRDHGWWKQHYEWRGNRWHPHPPPPPRHKKAPHDHGHGPGKHHR"
5564. .5668
/notes="REP (repetitive extragenic palindromic) element,
contains 2 REP sequences"
complement(5682..6458)
/genes="yaaA"
/locus_tag="S0006"
complement(5682..6458)
/genes="yaaA"
/locus_tag="S0006"
/notes="residues 1 to 258 of 258 are 100.00 pct identical
to residues 1 to 258 of 258 from Escherichia coli K-12 :
B0006"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAP15553.1"
/db_xref="GI:30039819"
/translation="MLIISPAKTLDYOSPLTTRTYTLPELDDNSQOLIHEARKLTPP
QISTLMRISDKLAGINAAFRHDMQDPFTANARQAILAFKGDVYTGLOAETSEDDFD
FAQQLHRLMSGLGYLVRPLDLDMQPYRLENGIRLENARGKDLYQFWGDIITNKLNEALA

```



[illegible]



Db	10449	GCCTTATCTCCCTCGCCTGAAATGCGCGCGCGAGGATTCAGTATC	10500
RESULT 11			
AP002550		281530 bp DNA linear BCT 07-MAR-2001	
LOCUS			
DEFINITION		Escherichia coli O157:H7 DNA, complete genome, section 1/20.	
ACCESSION		AP002550 BA000007	
VERSION		AP002550.1 GI:13359456	
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
A			

```

/translation="MKLYNLKDHNEQVSFAQAVTQGLKQKNOGLFFPHDLPEFLSLEID
EMKLDVFTSRAKILSAFIGEIPOEILEERVRAAFAPAPVANVESDVGCLELFGHP
TLAPKDFGFRMAQMLTHIAGDKPVTILTATSGDTGAAGAAHAFVGLPNVKVILYPEG
KISPLQKLEFCTIGGNETVALIDGDFDCAQALYKQAFDDDELKVALGINSANSINISR
LKAQICIFYFAVQAPQARLVVSPVSGNFDLTAGLLAKSLGLPKVKRPIAATNVN
DTPRFHLDGOWSPKATQATLSNAMDVSOPNNPRVEELFRRKIWLQKELGYAAVDDDE
TQOQTMRELKELGYTSPPHAAVYRALRDQLNPGFELFLGTAPAKFKESVFAILGE
TLDLPLKAEARADLPILLSHNLPAADFAALRLKLMNHQ"
gene
5251..5547
/genes="ECS0005"
CDS
5251..5547
/genes="ECS0005"
/notes="similar to B0005_ECOLI gi11786186 percent identity
95 in 98 aa (Conserved in E.coli K-12)"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="hypothetical protein"
/protein_id="BAB33428.1"
/db_xref="GI:13359461"
/translation="MKMQSIVLALSLLVAPMATAAEITLVPSVKLQIGDRDNRGY
YWDGHWDRDHGWKOHVYEWGRNRPHPPPPPRHKKHAKHDHGHGPKKHH"
gene
complement(5700..6476)
/genes="ECS0006"
CDS
complement(5700..6476)
/genes="ECS0006"
/notes="similar to YAAA_ECOLI gi11786187 percent identity
99 in 258 aa (Conserved in E.coli K-12)"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="hypothetical protein"
/protein_id="BAB33429.1"
/db_xref="GI:13359462"
/translation="MILLISPAKTLQXSLTTRYTLPELLNSQOLIHEARKLTPP
QISTLMRISKLGINAARHDQPDFTPENAKQAILAFKGDVYTGQAEFTFSEDDFD
FAQOHLRMLSLGLVRLDQMPYRLEMGIRLENARGKLDYQFWGDIITNKLNEALA
AQDENVVINLASDEYKSVKPKLNAAEIIKPFIDENKRGFKIISFYAKKARGILMSRF
IENRLTKPQLQGFNSGEYFFDFEASSNSGELVFKRYEQR"
gene
complement(6546..7976)
/genes="ECS0007"
CDS
complement(6546..7976)
/genes="ECS0007"
/notes="similar to YAAJ_ECOLI gi11786188 percent identity
99 in 476 aa (Conserved in E.coli K-12)"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="putative inner membrane transport protein"
/protein_id="BAB33430.1"
/db_xref="GI:13359463"
/translation="MPDFFSFINSVLGMSVMIYLLFGAGCWFTFRGTGFVQFVYIROFG
KSLKNSHPQPGGLTSFQSLCTSLAARVGSNLAGVALAITAGPGGAVFWWVAAFIG
MAYSFACSLAQLYKERDVGQPRGGPANTMARGLNRWGVLFVAPVLLIAYGIIIFSG
VQANAVARLSFDFPPLVTIGLIALFALLIIRLGHVARGLMQGVPLMALIWIWT
SLVTCVINGOLPHVINSIFESAFGQWGAAGAGYTLISQAITNGFORSFMSFNAAGM
STPNAAAASWHPHQAQGIOMIGIFIDTLVICTASAMLLIILAGNTGYMPLEGIQ
LLOKAMVLGMSGAEVPTLVILFAPFSIVANYIYAENNLFLRLNNKRAIWCRLIC
TFATVIGTLLSLPLMQDLADIIMACMAITNLITALLSPVHTIASDYLQRKLGR
VPVFDLPYDPIGRQLSRDANDVSYQE"
Query Match
Best Local Similarity 53.9%; Score 288; DB 1; Length 281530;
Matches 734; Conservative 0; Mismatches 605; Indels 33; Gaps 5;
QY 82 TCGTTGTCGGCTCGGCAAGAGCGGCGATGCCCTGTGGCGGTACCTGGCGGCGCGCGCT 141
DB 101720 TCATTATCGGCTGGCGCTCACCGGGCTTTCTCGGTGGAATTTTCTCGCTCGCGGTG 101779
QY 142 TGCCTTTCCGCGTGGTATACCGAGAACCGCGGAGCTGGCCACCTCGGTGCCGCC 201
DB 101780 TGACGCCGCGGCTTATGATACGCGTATGACACCGCTGGCTGGATAAATTACCGG --- 101836

```

```

202 AGTATCCGAGTGAAGTGGTTGGCGGCAACTCGACGCGAGTTCCTGCTGCGGCC 261
101837 -----AAGCGGTAGACGCCACACGGCGACTCTGAATGATGAATGCTGATGCGGCGAG 101890
262 GCGAACTCTATGTACGCCCCCGGCTTGTGCTGCGCACCCCCCTGCGCTGGTACAGGCCGCGG 321
101891 ATCTGATTTGCGCCAGTCCCGGTATTGCACTGGCGCATCCATCTTAAAGCCGCTGCGCGTG 101950
322 CGAAGCGGTGCGCATCTCCGCTGACATCGATCTTCTGCGCCGCGGAGCGAGCCCGCA 381
101951 ATGCCGGAATGAAATCGTTGGCGATATCGAGCTGTCTGTGCGGAGCACAAGCACCGA 102010
382 TCGTCGCCATCACCGGTTTCCAACGCGAAGACACCGTGCACACCTCTGTTGGCGCAATGG 441
102011 TTGTGCGGATTAACCGGTTCTAACGCGCAAAACACCGGTACACCTAGTGTGAATGG 102070
442 CGTGGCGCGGACAAAGGTGTCGCGTGGCGGCAACCTCGGCAACCCCGCGCTCGACC 501
102071 CGAAGCGCGGGGGTTAACGTTGTTGGTGGTGGCAATATTGGCTTGCCTGCGCTGTGATGC 102130
502 TGTGCGCGGACGACATCGAGCTGTACGTTGTTGGAGCTGTGAGCTTCCAGCTGGAACCT 561
102131 TGCTGGATGATGAGTGGAACTGTACGTACTGGAACCTGTGAGCTTCCAGCTGGAACCA 102190
562 GCGATCGCCTCAACCGCGAGGTGCGGACCGCTGCTGAACGTGACGCAAGACCATATGGATC 621
102191 CCTCCAGCTTACAGCGGTAGCAGCACCACTTCTGAACGTGACTGAAGATCATATGGATC 102250
622 GCTA---CGACGGCATGGCTGACTACACCTGGCGCAAGCAGCGGATCTTCCGCGGTGCC 678
102251 GCTATCGCTTGTGTTTACGCAATACCGTGGCGCAAAATTCGCTATTATTACAAAATGCGA 102310
679 GCCAGGTCGTGGTGAATCGCGCGATGCCCTGACCGGACCGCTGATCCCGCATACCGTGC 738
102311 AAGTTTGTGCTCAATGCTGATGCTTAAACATGCCCATTCGCGGTGCGGATGAC 102370
739 CGTCTGCTGCTTCCGCGCTTAACAAAGCGGACTTCAAGGCTTTCGCGCTGATCGAGGAAG 798
102371 GCTCGCTGAGCTTTGGCGTCAACATGGGTGACTATCA-----CCTGAATCATCAGC 102421
799 ACGGCCAGAGTGGCTGGCGTTCAGTTCGACAGCTGCTGCCGGTGGCGCACTGAGA 858
102422 AGGCGCAAACTGCTCGGCTGAAAGCGGAGAAAGTCTGAACGTGAAGAAATGAAC 102481
859 TCCGTGCGCGCCACCACTATTCCAAACGCGCTGCGCGCTGGCGTGGCGCATCGGCTGC 918
102482 TTTCCGGCGACATAACTACACCAATGGCTGGCGGCTGGCGTGGCAGATGCCGAG 102541
919 GCCTGCCGCTTCGACGCCATGCTCGCGCGCTGAAAGGGGTTTTTCCGGCTGCTCATCGCT 978
102542 GGTTCACCGGTGCCAGCAGCTGAAAGCGTTAAACACATTCACCTGCTGCGCGATCGCT 102601
979 GCCAGTGGGTACGCGAGCGGCGGCTGAGCTACTAGCAGATTCGAAGCGCCACCAAGC 1038
102602 TTGAAGTTGCTGGAGCATTAACCGGTACCTTGGATTAAAGTTCGAAGCGCACCAAGC 102661
1039 TCGCGCGCGCGCTCGCGCATCAGAGGCTGGGTGCGGACATCGACGCGCAAGTGGTGC 1098
102662 TCGCAGTACGGAAGCGCGCTGAATGCGCTG-----CAGTAGACGCGCAGCTGCATT 102715
1099 TGCTCGCGCGGAGAGCGCAAGGGCGCGATTTCCATGACCTGCGGAGCGGCTGCGGC 1158
102716 TGTTCGCGGCGGATGTTAAATCGCGGACTTTAGCCCACTTACCTGCTTACCTGAATG 102775
1159 GCTTCTCGCGGCGGTGTTACTGCTGGCGGTGACCGCGGCTGATTTCCCGCAGGCACTGG 1218
102776 GCGATAACGTACGTCTCTATTGTTTCGGTCTGTGACGCGGTGCACTGCGCGGTAC--- 102832
1219 GCAACGCGGTACCGCTGGTGGCGCTGCGCAACGCTGAGACAGCAGTCCGCGAGCGCGCG 1278
102833 ---GCCCGGAAGTGGCAGAACACCGAACAATGTAACACGCGATGCGCTTGTGCGCTC 102889
1279 AGCTGGCGCGGAGAGCGGATGCGGTGCTGTTGTGCGCGGCTGCGCGAGCGCTGGACATGT 1338

```

```

Db      102890  CGCGTGTTCAGCGCGGATATGGTTCTACTCTCCCGCGCTGTGCCAGCCTTGATCAGT 102949
QY      1339    TCAAGAACTTCGAAGAACCGGACGCGCTGTTCGCCAAAGCCGAGAGGAGCTAGCGTGAT 1398
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      102950  TCAAGAACTTCGAAGAACCGGACGCGCTGTTCGCCAAAGCCGAGAGGAGCTAGCGTGAT 103009
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY      1399    GCTGTCGGTGTTCGCGCCCTTCCCGTCGCCGCTGTTGAGCGGCACGGCATC 1450
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      103010  GCGTTTATCTCTCCCTCGCTGAAATGCGCGCGCTGCCAGGATTCAGTATC 103061
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 12
AE016755
LOCUS   Escherichia coli CFT073          300409 bp    DNA    linear    BCT 24-DEC-2002
DEFINITION  Escherichia coli CFT073 section 1 of 18 of the complete genome.
ACCESSION  AE016755 AE014075
VERSION    AE016755.1 GI:26106314
KEYWORDS
SOURCE    Escherichia coli CFT073
ORGANISM  Escherichia coli CFT073
          Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
          Enterobacteriaceae; Escherichia.
REFERENCE  1 (bases 1 to 300409)
          Welch,R.A., Burland,V., Plunkett,G.D. III, Redford,P., Roesch,P.,
          Rasko,D.A., Buckles,E.L., Liou,S.-R., Boutin,A., Hackett,J.,
          Stroud,D., Mayhew,G.F., Rose,D.J., Zhou,S., Schwartz,D.C.,
          Perna,N.T., Mobley,H.L.T., Donnenberg,M.S. and Blattner,F.R.
          Extensive Mosaic Structure Revealed by the Complete Genome Sequence
          of Uropathogenic Escherichia coli
          Proc. Natl. Acad. Sci. U.S.A. 99 (26), 17020-17024 (2002)
12471157
REFERENCE  2 (bases 1 to 300409)
          Welch,R.A., Burland,V., Plunkett,G.D. III, Redford,P., Roesch,P.,
          Rasko,D.A., Buckles,E.L., Liou,S.-R., Boutin,A., Hackett,J.,
          Stroud,D., Mayhew,G.F., Rose,D.J., Zhou,S., Schwartz,D.C.,
          Perna,N.T., Mobley,H.L.T., Donnenberg,M.S. and Blattner,F.R.
          Direct Submission
          Submitted (20-JUN-2002) Genetics Laboratory, University of
          Wisconsin - Madison, 445 Henry Mall, Madison, WI 53706, USA
          Location/Qualifiers
            1..300409
              /organism="Escherichia coli CFT073"
              /mol_type="genomic DNA"
              /strain="CFT073"
              /db_xref="taxon:199310"
            190..255
              /gene="thrL"
              /locus_tag="c5491"
            190..255
              /gene="thrL"
              /locus_tag="c5491"
            190..255
              /gene="thrL"
              /locus_tag="c5491"
              /function="leader; Amino acid biosynthesis: Threonine"
              /note="thr operon attenuator; Escherichia coli k-12
              ortholog: b0001; Escherichia coli O157:H7 ortholog: z0001"
              /codon_start=1
              /transl_table=11
              /product="Thr operon leader peptide"
              /protein_id="AA078501.1"
              /db_xref="GI:26106315"
              /translation="MKRISTTTTITTTTGTGNGAG"
            308..1018
              /insertion_seq="IS1541A-like"
            534..911
              /locus_tag="c0002"
            534..911
              /locus_tag="c0002"
              /note="Residues 2 to 125 of 125 are 94.35 pct identical to
              residues 29 to 152 of 152 from SwissProt.40 :
              >sp|Q57334|T200_SALTY Transposase for insertion sequence
              element IS200"
              /codon_start=1
              /transl_table=11

```

```

/product="Hypothetical protein"
/protein_id="AA078502.1"
/db_xref="GI:26106316"
/translation="MFYREKRAAGCILKRLCEWKSVRILEAECCADHIMLVETPPK
MSVGFMCYLVKSKSLMPYEOFGDLKFKYRNREFWCRGYVVDTVGKNTAKIQDIYIKHQ
LEEDKMGEOGLSPYPGSPFTGRK"
985..3510
/gene="thrA"
/locus_tag="c0003"
985..3510
/gene="thrA"
/locus_tag="c0003"
/function="enzyme; Amino acid biosynthesis: Threonine"
/note="Bifunctional; Escherichia coli K-12 ortholog:
b0002; Escherichia coli O157:H7 ortholog: z0002;
bifunctional; Homoserine dehydrogenase I"
/codon_start=1
/transl_table=11
/product="Aspartokinase I"
/protein_id="AA078503.1"
/db_xref="GI:26106317"
/translation="MKNRLCRRIFTSTKGNVTTMRVLKFGSTSVANAERFLRVADI
LESNARQGVATVLSAPAKITNHLVAMIEKTTISGGDALPNISDAERIPAEILLTGLAA
QGFPLAQLKTFVDEFFAQIKHVLHGHSLLGQCPDSINAAICRGKEMSIATMAGVLE
ARGHNVITDVPVEKLLAVGHYLESTVTAESTRTTAASRIADHVMVAGFTAGNEKG
ELSVLRGNSDYSAVALAACLRADCEIWDVGYTCDPQVQPARLLKMSYQIGAM
ELSYFGKVLHPRTTTPIAQFOICLINTGQPQAPGTIGASRDEDELPPKVSIGNLN
NAMPSVSGPMGKMGVMAARFAAMSRARISVLLITOSSSEYSISFCVPOSDCVRAE
RAMQSEFLELEKGLLEPLATVERLAISVVGDMRTLRIGISAKFAALANINIVA
IAQGSSEISISVNNDDATGTVTHIISVQKMLFNIDQIEVFYIGVGVGALLDEKLRQA
QSMKNKHIDLRVCGVANSKALLTNVHGLNLENWQELAAQAKPEPNLRLIRLVKEYH
LINPVIVDCTSSQAVDQYADFREGFHVVTNNKANTSSMDYYHQLRYAAEKSRKE
EMGYTEPPDDLSGMDVARKLLILARETGRELEADIEIEPVLPAEFNAEGDVAEFM
ANLSQDLNFAARVAKARDECKVLYVGNIDEDGVCVRKIAEVDSDNPLFKVKNGENA
LAFYSHYQPLPLVLRGAGNDVTAAGVAFDRLTLRWKLCGV"
3512..4444
/gene="thrB"
/locus_tag="c0004"
3512..4444
/gene="thrB"
/locus_tag="c0004"
/EC_number="2.7.1.39"
/function="enzyme; Amino acid biosynthesis: Threonine"
/note="Escherichia coli K-12 ortholog: b0003; Escherichia
coli O157:H7 ortholog: z0003"
/codon_start=1
/transl_table=11
/product="Homoserine kinase"
/protein_id="AA078504.1"
/db_xref="GI:26106318"
/translation="MVKVYAPASSANMSVGFDLGAATVPVDGALLGDVVTVEAAETF
SLNNRGPADKLPSRENIVYQCWERECQELGKOIPVAMTLEKNMPTGGSGIGSSACS
VVAALMANNEHCCKPLNDTRLLALMGELEGRSSGSIHDVNPAPCLGQMQLNENDI
ISQQVPGDEMDWLVAIPGKIVSTAEARAILPAQRQDCIANGHRTGAFIHACYSRQ
PELANMLMDVIAEYPERLLPFGFQAQVAEIAIGAVASGISGSGPTFALCDKPDPTA
QRVADWLGNLYLQNOEGFVHCIRLDTAGARVLEN"
4445..5731
/gene="thrC"
/locus_tag="c0005"
4445..5731
/gene="thrC"
/locus_tag="c0005"
/EC_number="4.2.3.1"
/function="enzyme; Amino acid biosynthesis: Threonine"
/note="Escherichia coli K-12 ortholog: b0004; Escherichia
coli O157:H7 ortholog: z0004"
/codon_start=1
/transl_table=11
/product="Threonine synthase"
/protein_id="AA078505.1"
/db_xref="GI:26106319"
/translation="MKLINLKDHNQVSVFAQAVTQGLGKNQGLFFPHDLPFSLTEID

```



gene	5799..5948	/locus_tag="c0006"	EMLKDFVTRSAKILSAFIDGPIQEIIEERVRAAFAPAPVANVEDVGCLEFHFG
CDS	5799..5948	/locus_tag="c0006"	TLAPKGGRRMAQILTHIAGDKPVTILTATSGDTGAFAVAHAFYGLPNVKVILYIPRG
		/codon_start=1	KISPKQELKFTLCGGNTETVAIDGDFDACQALVKQAFDDDEELKVALGNSANGINISR
		/transl_table=11	LLAQYIFEAVALQPGEARNLQVSVPSGNFGDLTAGLLAKSLGLPVKRFRTAATNVN
		/product="Hypothetical protein"	DTQPRFLHDGQSPKATQAVLNSNADYVQPNPNRVEELFRRKIWLQKELGYAAVDDE
		/protein_id="AAW78506.1"	TTQOTRELKELGVTSPHAHAYRALRDOLNPCEYGLGTGTAHPAKFKESVEAILGE
		/db_xref="GI:26106320"	TLDPKLELAERADLPLLSHNLPADEAALRKLMMNHQ
		/translation="MEANDREKGEILKCGNLEIRIAENNNCRSHRVISGYRPTGND	
gene	5945..6241	/locus_tag="c0007"	KRSNL
CDS	5945..6241	/locus_tag="c0007"	5945..6241
		/codon_start=1	/locus_tag="c0007"
		/transl_table=11	/note="Escherichia coli K-12 ortholog: b0005; Escherichia
		/product="Hypothetical protein yaaX precursor"	coli O157:H7 ortholog: z0005"
		/protein_id="AAW78507.1"	/codon_start=1
		/db_xref="GI:26106321"	/transl_table=11
		/translation="MKWQMSIVIALSLVLVAPMAQAAEITLVPSVKLQIGDRDNRGY	/product="Hypothetical protein"
		YWDGHRDRHGWRKQHYEGRWGNRWHYPGPPSPRNKHNDRGDRHPGPKHHR"	/protein_id="AAW78508.1"
		complement(6021..6269)	/db_xref="GI:26106322"
		/locus_tag="c0008"	/translation="MPDGNPALIFSDVDVCAOPDDRDHDDHCACYGAAVAHVHGASDC
		complement(6021..6269)	LAINVYSTSRGRASDRHPSNSHDYHDLRSVILLTGLA"
		/locus_tag="c0008"	6060..6323
		/codon_start=1	/locus_tag="c0009"
		/transl_table=11	6060...6323
		/product="Hypothetical protein"	/locus_tag="c0009"
		/protein_id="AAW78509.1"	/codon_start=1
		/db_xref="GI:26106323"	/transl_table=11
		/translation="MIIVATIGMAVTGATTAGGNNIMNGEAIAGTHMDRRHRAITST	/product="Hypothetical protein"
		MIIVATIGRTNIARIQYQRIARHYPALSGIDYKLVNVRVSV"	/protein_id="AAW78509.1"
		complement(6301..7077)	/db_xref="GI:26106323"
		/gene="yaaX"	/translation="MIIVATIGMAVTGATTAGGNNIMNGEAIAGTHMDRRHRAITST
		/locus_tag="c0010"	MIIVATIGRTNIARIQYQRIARHYPALSGIDYKLVNVRVSV"
		complement(6301..7077)	/db_xref="GI:26106324"
		/gene="yaaX"	/translation="MIIVATIGMAVTGATTAGGNNIMNGEAIAGTHMDRRHRAITST
		/locus_tag="c0010"	MIIVATIGRTNIARIQYQRIARHYPALSGIDYKLVNVRVSV"
		complement(6301..7077)	MIIVATIGRTNIARIQYQRIARHYPALSGIDYKLVNVRVSV"
		/gene="yaaX"	MIIVATIGRTNIARIQYQRIARHYPALSGIDYKLVNVRVSV"
		/locus_tag="c0010"	MIIVATIGRTNIARIQYQRIARHYPALSGIDYKLVNVRVSV"
		complement(6301..7077)	MIIVATIGRTNIARIQYQRIARHYPALSGIDYKLVNVRVSV"
		/gene="yaaX"	MIIVATIGRTNIARIQYQRIARHYPALSGIDYKLVNVRVSV"
		/locus_tag="c0010"	MIIVATIGRTNIARIQYQRIARHYPALSGIDYKLVNVRVSV"
		complement(6301..7077)	MIIVATIGRTNIARIQYQRIARHYPALSGIDYKLVNVRVSV"
		/gene="yaaX"	MIIVATIGRTNIARIQYQRIARHYPALSGIDYKLVNVRVSV"
		/locus_tag="c0010"	MIIVATIGRTNIARIQYQRIARHYPALSGIDYKLVNVRVSV"
		complement(6301..7077)	MIIVATIGRTNIARIQYQRIARHYPALSGIDYKLVNVRVSV"
		/gene="yaaX"	MIIVATIGRTNIARIQYQRIARHYPALSGIDYKLVNVRVSV"
		/locus_tag="c0010"	MIIVATIGRTNIARIQYQRIARHYPALSGIDYKLVNVRVSV"
		complement(6301..7077)	MIIVATIGRTNIARIQYQRIARHYPALSGIDYKLVNVRVSV"
		/gene="yaaX"	MIIVATIGRTNIARIQYQRIARHYPALSGIDYKLVNVRVSV"
		/locus_tag="c0010"	MIIVATIGRTNIARIQYQRIARHYPALSGIDYKLVNVRVSV"
		complement(6301..7077)	MIIVATIGRTNIARIQYQRIARHYPALSGIDYKLVNVRVSV"
		/gene="yaaX"	MIIVATIGRTNIARIQYQRIARHYPALSGIDYKLVNVRVSV"
		/locus_tag="c0010"	MIIVATIGRTNIARIQYQRIARHYPALSGIDYKLVNVRVSV"
		complement(6301..7077)	MIIVATIGRTNIARIQYQRIARHYPALSGIDYKLVNVRVSV"
		/gene="yaaX"	MIIVATIGRTNIARIQYQRIARHYPALSGIDYKLVNVRVSV"
		/locus_tag="c0010"	MIIVATIGRTNIARIQYQRIARHYPALSGIDYKLVNVRVSV"
		complement(6301..7077)	MIIVATIGRTNIARIQYQRIARHYPALSGIDYKLVNVRVSV"
		/gene="yaaX"	MIIVATIGRTNIARIQYQRIARHYPALSGIDYKLVNVRVSV"
		/locus_tag="c0010"	MIIVATIGRTNIARIQYQRIARHYPALSGIDYKLVNVRVSV"
		complement(6301..7077)	MIIVATIGRTNIARIQYQRIARHYPALSGIDYKLVNVRVSV"
		/gene="yaaX"	MIIVATIGRTNIARIQYQRIARHYPALSGIDYKLVNVRVSV"
		/locus_tag="c0010"	MIIVATIGRTNIARIQYQRIARHYPALSGIDYKLVNVRVSV"
		complement(6301..7077)	MIIVATIGRTNIARIQYQRIARHYPALSGIDYKLVNVRVSV"
		/gene="yaaX"	MIIVATIGRTNIARIQYQRIARHYPALSGIDYKLVNVRVSV"
		/locus_tag="c0010"	MIIVATIGRTNIARIQYQRIARHYPALSGIDYKLVNVRVSV"
		complement(6301..7077)	MIIVATIGRTNIARIQYQRIARHYPALSGIDYKLVNVRVSV"
		/gene="yaaX"	MIIVATIGRTNIARIQYQRIARHYPALSGIDYKLVNVRVSV"
		/locus_tag="c0010"	MIIVATIGRTNIARIQYQRIARHYPALSGIDYKLVNVRVSV"
		complement(6301..7077)	MIIVATIGRTNIARIQYQRIARHYPALSGIDYKLVNVRVSV"
		/gene="yaaX"	MIIVATIGRTNIARIQYQRIARHYPALSGIDYKLVNVRVSV"
		/locus_tag="c0010"	MIIVATIGRTNIARIQYQRIARHYPALSGIDYKLVNVRVSV"
		complement(6301..7077)	MIIVATIGRTNIARIQYQRIARHYPALSGIDYKLVNVRVSV"
		/gene="yaaX"	MIIVATIGRTNIARIQYQRIARHYPALSGIDYKLVNVRVSV"
		/locus_tag="c0010"	MIIVATIGRTNIARIQYQRIARHYPALSGIDYKLVNVRVSV"
		complement(6301..7077)	MIIVATIGRTNIARIQYQRIARHYPALSGIDYKLVNVRVSV"
		/gene="	

Query Match 19.8%; Score 285.4; DB 1; Length 300409;  
Best Local Similarity 54.0%; Pred. No. 1.1e-25;  
Matches 715; Conservative 0; Mismatches 576; Indels 33; Gaps 5;

Qy	80	CATCGTTGTCGGCTCCGCAAGAGCGGCGATCTCCCTGTCGCTACCTTGGCGCGCCGCG	139
Db	100940	CGTCATTATCGCCTGGCGCTCACCGGCTTCTTGCCTGGACTTTTCTCCTCGCTCGCG	100999
Qy	140	CTTGCCCTTCCCGTGGTGCATACCCGAGAACCCCGGAGCTGGCCACCCTCGGTGC	199
Db	101000	TGTGACGCGCGCTTATGAGATACGCGTATGACACCGCTGGCCTTGGATTAATTAACCG-	101058
Qy	200	CCAGTATCCGAGGTGGAAGTGCCTTGGCGGAACTGCACGCGGAGTTCCTCTGCTCGC	259
Db	101059	-----AAGCCGTAGAACGCCACACGCGCGTCTGAATGATGAATGCTGATGGCGC	101110
Qy	260	CCGCGAATCTATGTACGCCCGGCTTGTGCTGGCACCCCTGCGCTGGTACAGCCGCG	319
Db	101111	AGATCTGATTTCGCCAGCTCCGCTATTGCACTGGCGCATCCATCTTAAGCGCTCGCG	101170
Qy	320	CGCGAAAGGCGTGCATCTCCGGTGACATCGATCTTTCGCCGCGAGGCGAAGCCGCC	379
Db	101171	TGATGCCGGAATCGAAATCGTTGGCGATATCGAGCTGTCTGCTGGAAGCAACAGCAC	101231
Qy	380	GATCGTCGCCATCATCCGGTTCACCGCAAGACACCGTGACCACTTGGTGGGCGAAT	439
Db	101231	GATTGTGCGATTAACCGGTTCTAACGCGAAAAACGCTCACACGCTAGTGGGTGAAT	101290
Qy	440	GCGCGTGGCCCGGACAAAGCGTGTCCCGTGGCGGCAACCTGGCACCCCGCGCTCGA	499
Db	101291	GGCGAAAGCGCGGGGTAAAGTGTGGTGGCGCAATATTGGCCCTGCCCTGCTGAT	101350
Qy	500	CTGCTGGCCGACGACATCGAGCTGTACGTGTGGAGCTGTGAGCTTCCAGCTCGAAT	559
Db	101351	GCTGTGGATGATGATGCGAACTGTAGCTGTGGAATGTGAGCTTCCAGCTGGAAC	101410
Qy	560	CTGCGATCGCCTCAACGCGAGGTGGCAGCGTGTGAACGTACGCAAGACCATATGA	619
Db	101411	CACCTCCAGCTTACAGGCGGTAGCAGCGACCATTTCTGAACGTGACTGAAGATCATATGA	101470
Qy	620	TCGCTA---CGAAGGCTGGTGTGACTACCACTGGCCAAAGCACCGGATCTCCGCGGTGC	676
Db	101471	TCGCTATCCGTTGTGTTTACAAACAGTATCGTGCAGCAAACTTCGCATTTACGAAAACG	101530
Qy	677	CCGCGAGTCTGTGTGAATCGCGGATGCCCTGACCGCGGTGATCGCGGATACGCT	736
Db	101531	AAAAGTTGCGTGGTTAATGCTGATGATGCTTTAAATGCGGATTCGGGTGCGGATGA	101590
Qy	737	GCGCTGTGCTGTTCGGCTTGAACAAGCCGAGCTTCAAGGCTTTCGGCCTGATCGAGGA	796
Db	101591	ACGCTGCTGACGTTTGGCGTCAACATGGCGACTATCA-----CCTGAATCATCA	101641
Qy	797	AGACGGCCAGAAGTGGCTGGCGTTCCAGTTCGACAAGCTGTGCGCGTTGGCGAATGAA	856
Db	101642	CGAGGGCGAAACCTGGCTGCGGGTGAAGGGCGAAGAGTGTGAACGTGAAGAATGAA	101701
Qy	857	GATCGTGGCGCCACAACTATCCAACGCGCTGCGCGCTGGCGCTGGCCATGCGGT	916
Db	101702	ACTTTCGGGCGAGATAACTACCAATGCGCTGGCGCGTGGCGCTGGCAGATGCGCG	101761
Qy	917	CGGCTGCGGTCGACGCCATGCTCGGCGCGTGAAGCGTTTTTCGGCGCTGGCTCATCG	976
Db	101762	AGGGTACCGGCTGCCAGCACCTTGAAGCGTTAACCATTCATCTGGTCTGCCGATCG	101821
Qy	977	CTGCCAGTGGTACGCGAGCGGCGGTGAGCTTACGACGATTCGAAGCGCACCAA	1036
Db	101822	CTTTGAAGTTGCTGGAGCATAACGGTGTGCGCTGGGTTAACGATTTAAAAGCGACCAA	101881
Qy	1037	CGTCGGCGCGCCTTGGCGGCGATCGAGGGGCTGGTGCCGACATCGAGCGGAAGCTGTT	1096
Db	101882	CGTCGGAGTACGGAAGCGCGCTGAATGGCTGTCATG-----TAGACGCGACGCTGA	101935
Qy	1097	GCTCTCGCGCGGAGACGCAAGGGCGCGATTTCCATGACCTGCGCGAGCGCGTGC	1156
Db	101936	TTTGTGCTGGGTGGCATGTAAATCGCGGACTTTTAGCCCACTGGCGGCTTACCTGAA	101995
Qy	1157	GCGCTTCTGCGCGGCTGTGCTGTTGGCGGTGACGCGCGGCTGATTTGCCAGGCACT	1216

Db 101996 TGCGATTAACGTACGTGTATTGTTTCGGTCGTGACGGCGCAGCTGGCGCGTAC- 102054  
 QY 1217 GGGCAACGCGGTACCGTGTGCGCGTCAACAGCTGACGAGCAAGCACTGGCGAGCCGC 1276  
 Db 102055 -----GCCCGAGTGCAGCAACAAACCGAACTATGGAACAGCGATGCGCTGTGCTGC 102109  
 QY 1277 CGAGCTGGCCCGCAGGCGATGCGGTGTGTTGTCGCCGCGCTGCGAGCCCTGGACAT 1336  
 Db 102110 TACGCGTGTTCAGCGCGCATATGTTGCTCTCCCGCGCTGTGCCAGCCTGTATCA 102169  
 QY 1337 GTTCAAGAACTTGAAGAACGCGGACCGCTGTTGCGCAAAAGCCGTAGAGAGCTAGCGTG 1396  
 Db 102170 ATTCAGAACTTGAAGAACGCGGACCGATGAGTTTGCCCGCTGTCGCGAAGGAGTAGTTG 102229  
 QY 1397 ATGC 1400  
 Db 102230 ATGC 102233

## RESULT 13

## ECMUROY

## DEFINITION

## ACCESSION

## VERSION

## KEYWORDS

## SOURCE

## ORGANISM

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## MEDLINE

## PUBMED

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## COMMENT

## FEATURES

## source

## RBS

## CDS

TGNNFASYLHPIPLRHAGELIVCTAIVGAGLFLWFNTYPAQVFMGDVGSALGGA  
 LGTIAVLRQEFLLVIMGVVYVETLSVLQVGSFKLRGQIRFMAPIHIIHYELKGNP  
 EPVIVREVIISLMLVLIGLATLVR"  
 1018..1023  
 /note="pot. -35 region"  
 1043..1048  
 /note="pot. -10 region"  
 1181..1185  
 /note="pot. ribosome binding site"  
 1196..2512  
 /note="murd gene product (AA 1-438)"  
 /codon\_start=1  
 /transl\_table=11  
 /protein\_id="CAA35933.1"  
 /db\_xref="GI:42058"  
 /db\_xref="SWISS-PROT:P14900"  
 /translation="MADYQGNKVVIIIGLTLGLSCVDFFLIARGVTRVMDTRMTPPGL  
 DKLPEAVERHTGSLNDEWLMADLIIVASPGIALAHPSLSAADAGIEVGDIEFCRE  
 AQAPVITGTSNGKSTVTLVGEAKAAGVNVGVGNIGLIPALMLLDCELYLELS  
 SFOLETTSSLOAAATILNVTEHMDRYPGLQOYRAAKLRIVENAKVCVVNADALT  
 MPTRGADERCVSFGVNNQDYHLNHOGETWLVYKGEKLVNVEKLSGOHNTNALAA  
 LALADAGLPRASSLALATFTGLPHRFVVLVLEHNGVWINDSKATNRVSGTEAANGL  
 HVDGTHLLLGSGKADSPFLARYLNGVNRVLYCFGRDGAALALRPEVAEGTETME  
 QAMRLIAPRVQGDVWLLSPACASLDQFKNFQGRNEFARLAKELG"  
 BASE COUNT 530 a 631 c 777 g 670 t  
 ORIGIN  
 Query Match 19.4%; Score 282; DB 1; Length 2608;  
 Best Local Similarity 53.2%; Pred. No. 1.1e-24;  
 Matches 731; Conservative 0; Mismatches 610; Indels 33; Gaps 5;  
 QY 80 CATCGTTGTCGGCTCGGCAAGAGCGGCATGTCCCTGGTGCCTACCTGGCGCGCGCGG 139  
 Db 1222 CGTCATTATCGGCTGGCGCTCACCGGGCTTTCTCGTGGACTTTTCTCTCGCTCGCG 1281  
 QY 140 CTTGCGCTTTCGGCGTGTGATACCGAGAGAACCGCGCGAGCTGCCACCTCGGTGTC 199  
 Db 1282 TGTGACGCGCGCGCTTATGATACGCGTATGACACCGCTTGGCTGGATTAATACCGG- 1340  
 QY 200 CCAGTATCGCAGGTGGAAGTGTGTCGGCGCAACTCGACGCGGAGTTCCTCTGCTCCGC 259  
 Db 1341 -----AAGCGTAGAACGCCACACGGGCAGTCTGATGATGATGGCTGATGGCGG 1392  
 QY 260 CCAGCAACTCTATGTGACGCGCGCGGTGTGCTGCGCAACCCCTGCGCTGTGACAGCGCG 319  
 Db 1393 AGATCTGATGTGCGCAGTCCCGGTATGACACTGGCGCATCCATCTTAAGCGCTGCGCG 1452  
 QY 320 CGCGAAGGCGTGGCATCTCCGCTGACATCGATCTTTCGCCCGCGAGCGGAAGCCCC 379  
 Db 1453 TGATGCCGAATCGAAATCGTTGGCGATATCGAGCTGTCTGTCGGAAGCACAAGCACC 1512  
 QY 380 GATCGTGGCATCACCGGTTCACCAACGGAAGAGCAGCCGTGACCAACCCCTGGTGGCGAAAT 439  
 Db 1513 GATTGTGGCGATTACCGGTCTTAACGCGCAAAACAGCGGTACCACGCTAGTGGTGAAT 1572  
 QY 440 GCGGCTGGCGCGGACAAAGCGTGTCCCGGTGCGGCGAAACCTCGGCAACCCCGCGGTGCA 499  
 Db 1573 GCGCAAGCGCGCGGGGTAAACGTTGCTGTGGTGGCAATATTGGCTGCTCGCTGAT 1632  
 QY 500 CCGTGGTGGCGGACGACATCGAGCTGTACGTGTGGAGCTGTCCAGCTTCCAGCTGGAAAC 559  
 Db 1633 GCTACTGGATGATGAGTGTGAACGTGTGCGTGGAACTGTGCGAGTTCAGCTGGAAAC 1692  
 QY 560 CTGCGATCGCTCAACCGCGAGGTGGCGGCTGCTGAACGTGACGAGCAAGACCATATGA 619  
 Db 1693 CACCTCCAGCTTACAGCGGTGACGAGGACCATCTTGAACGTGACGTGAAGATCATATGA 1752  
 QY 620 TCGCTA---CGAGCGCATGGCTGACTACCACTGGCGCAAGACCGATCTTCGCGCGTGC 676  
 Db 1753 TCGCTATCGGTTGGTTTACAACAGTATCGTGCAGCAAAACCTGCGCATTTTACGAAACGC 1812  
 QY 677 CCGCCAGGTGCGGTGAATCGCGCGCATCCCTGACCCCGCGCTGATCGCCGATACCTG 736

```
Db 1813 GAAAGTTTCGCTGGTTAATGCTGATGATGCCCTTAACAATGCCGATTCGCGGTGGCGGATGA 1872
QY 737 GCCGTGCTGGTTCGCTGGCGTGAACAACCGAGACTTCAAGGCTTTTCGCGCTGATCGAGGA 796
Db 1873 ACGTGTGCTGAGTTCGCTGCAACATGGTGAATCTATCA-----CCTGAATCATCA 1923
QY 797 AGACGGCCAGAAAGTGGCTGGCGTTCAGTTCGCAAGCTGTCGCGGTGGCGCACTGAA 856
Db 1924 GCAGGGCGAAACCTGGCTGGCGGTAAAGCGGAGAAAGTGTGAATGTGAAAGAGATGAA 1983
QY 857 GATCGGTGGCGCCACACATATTCCACGCGCTGCGCGCTGGCGTGGCGTGGCGGCTGCGGT 916
Db 1984 ACTTCCGGGAGCAGTACTACTACCAATGCTGCGTGGCGCTGGCGTGGCGAGTGTGC 2043
QY 917 CGGCTGCGCTTCGACGCCATTCGTCGGCGCGCTGAAGCGCTTTCCCGCGCTGGCTCAFCG 976
Db 2044 AGGTTTACCGGTGCCAGCAGCCTGAAGCGCTTAACCAATCACTGCTGCTGCCGATCG 2103
QY 977 CTGCCAGTGGTACGCGAGCGGAGCGGTGAGTCTACTACACCATTCACAGGCCACCA 1036
Db 2104 CTTTGAAGTTGTGTGGAGCATAACGCGGTACGTTGGATTAACTGAAAGTGAAGGACCA 2163
QY 1037 CGTCGGCGCCCTCGCGCGATCGAGGGCTGGTGGCGACATCGACGGCAAGCTGGT 1096
Db 2164 CGTCGGCAGTACGGAACGGCGTGAATGGCTG-----CAGTAGACGCACACTGCA 2217
QY 1097 GCTGCTCGCGCGGAGACGCAAGGCGCGCATTTCCATGACCTGCGCGAGCGCGTGC 1156
Db 2218 TTTTGTGCTGGTGGCGATGTAATCGCGGACTTTAGCCCACTGGCGCGTTACCTGAA 2277
QY 1157 GCGCTTTCGCGCGGTGTTGCTGCTGGCGGTGACCGCGGCTGATTCGCCAGGCACT 1216
Db 2278 TGGCGATAACACTAGCTCTGATTGTTGCGTGTGACGCGCGAGCTGGCGGCGTAC- 2336
QY 1217 GGGCAACGCGTACGCTGGTGGCGTGCACACCTGGACGAAGCAAGTCCGCGAGCGCG 1276
Db 2337 -----GCCGGAATGGCAGAACAAACGAACTATGGAACAGCGGATGCCCTGTGCTGGC 2391
QY 1277 CGAGCTGGCGCGGAGCGGATGCTGCTGTTGCGCGGCTTCGCGAGCCTGGACAT 1336
Db 2392 TCCGCTGTTTCAGCGGGCGATATGTTCTGCTCTCCCGACGCTGTGCCAGCCTTGATCA 2451
QY 1337 GTTCAAGACTTCAAGAACGCGAGCGCTGTTGCGCAAGCCGTAGAGAGTAGCGTG 1396
Db 2452 GTTCAAGAACTTTGAACAACGAGCAATGAGTTTGGCCGCTGTCGCGAAGGAGTTAGTTG 2511
QY 1397 ATGCTGTGCGTGTGGCGCCCTTCCCGTGGCGGCTGTTGAGCGCGGACGGCATC 1450
Db 2512 ATGCGTTTATCTCTCCCTCGCTGAAATGCCGCGCTGCCAGGATTCAGTATC 2565

RESULT 14
AE000118 Escherichia coli K12 M1655 section 8 of 400 of the complete genome. linear BCT 01-DEC-2000
LOCUS AE000118.1 GI:1786262
ACCESSION AE000118 U00096
VERSION AE000118.1
KEYWORDS
SOURCE
ORGANISM Escherichia coli K12
Escherichia coli K12
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; Escherichia.
REFERENCE
Blattner, F.R., Plunkett, G. III, Bloch, C.A., Perna, N.T., Burland, V.,
Riley, M., Collado-Vides, J., Glasner, J.D., Rode, C.K., Mayhew, G.F.,
Gregor, J., Davis, N.W., Kirkpatrick, H.A., Goeden, M.A., Rose, D.J.,
Mau, B. and Shao, Y.
The complete genome sequence of Escherichia coli K-12
Science 277 (5331), 1453-1474 (1997)
MEDLINE 97426617
PUBMED 9278503
REFERENCE 2 (bases 1 to 21757)
```

AUTHORS  
TITLE  
JOURNAL

Blattner, F.R.  
Direct Submission  
Submitted (16-JAN-1997) Guy Plunkett III, Laboratory of Genetics,  
University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.  
Email: ecoligenetics.wisc.edu Phone: 608-262-2534 Fax:  
608-263-7459

REFERENCE  
AUTHORS  
TITLE  
JOURNAL

3 (bases 1 to 21757)  
Blattner, F.R.  
Direct Submission  
Submitted (02-SEP-1997) Guy Plunkett III, Laboratory of Genetics,  
University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.  
Email: ecoligenetics.wisc.edu Phone: 608-262-2534 Fax:  
608-263-7459

REFERENCE  
AUTHORS  
TITLE  
JOURNAL

4 (bases 1 to 21757)  
Blattner, F.R.  
Direct Submission  
Submitted (13-OCT-1998) Laboratory of Genetics, University of  
Wisconsin, 445 Henry Mall, Madison, WI 53706, USA

## COMMENT

This sequence was determined by the E. coli Genome Project at the  
University of Wisconsin-Madison (Frederick R. Blattner, director).  
Supported by NIH grants HG00301 and HG01428 (from the Human Genome  
Project and NCHGR). The entire sequence was independently  
determined from E. coli K12 strain M1655. Predicted open reading  
frames were determined using Genemark software, kindly supplied by  
Mark Borodovsky, Georgia Institute of Technology, Atlanta, GA,  
30332 [e-mail: mark@ember.gatech.edu]. Open reading frames that  
have been correlated with genetic loci are being annotated with CG  
Site Nos., unique ID nos. for the genes in the E. coli Genetic  
Stock Center (CGSC) database at Yale University, kindly supplied by  
Mary Berlyn. A public version of the database is accessible  
(http://cgsc.biology.yale.edu). Annotation of the genome is an  
ongoing task whose goal is to make the genome sequence more useful  
by correlating it with other data. Comments to the authors are  
appreciated. Updated information will be available at the E. coli  
Genome Project's World Wide Web site  
(http://www.genetics.wisc.edu). \*\*\* The E. coli K12 sequence and  
its annotations are periodically updated; this is version M54. No  
sequence changes. Annotation updates: updated gene identifications  
and products; all new functional assignments courtesy of Monica  
Riley; added promoters, protein binding sites, and repeated  
sequences described in reference 1. The unique numeric identifiers  
beginning with a lowercase 'b' assigned to each gene (protein- or  
RNA-encoding) are now designated as gene synonyms instead of  
labels. This should allow them to be searched for in Entrez as gene  
names.

FEATURES  
source

Location/Qualifiers  
1. 21757  
/organism="Escherichia coli K12"  
/mol\_type="genomic DNA"  
/strain="K12"  
/sub\_strain="M1655"  
/db\_xref="taxon:83333"  
complement(77..163)  
/gene="leuL"  
/note="synonym: b0075"  
complement(77..163)  
/gene="leuL"  
/function="leader; Amino acid biosynthesis: Leucine"  
/note="f28; 100 pct identical to LPL\_ECOLI SW: P09149"  
/codon\_start=1  
/product="leu operon leader peptide"  
/protein\_id="AAC73186.1"  
/db\_xref="GI:1786263"  
/translation="MTHIVRFIGILLNASSLRGRRVSGIOH"  
complement(197..226)  
/note="factor Sigma70; promoter leu; documented +1  
at 83735"

## gene

## CDS

## promoter

## protein\_bind

## protein\_bind

complement(201..221)  
/note="No predicted promoter"  
/bound\_moiety="LexA predicted site"  
complement(201..221)  
/note="central position to predicted promoter: -0.5"

```

/bound_moiety="LexA predicted site"
complement(354..368)
/notes="No predicted promoter"
/bound_moiety="Lrp documented site"
complement(354..368)
/notes="central position to leuABCD promoter: -156.5"
/bound_moiety="Lrp documented site"
460..474
/notes="central position to predicted promoter: -28.5"
/bound_moiety="Lrp predicted site"
462..490
/notes="factor Sigma70; predicted +1 start at 84042"
complement(519..537)
/notes="central position to predicted promoter: -319.5"
/bound_moiety="tyrR predicted site"
554..565
/notes="central position to predicted promoter:64"
/bound_moiety="AraC predicted site"
646..1767
/genes="leuO"
/notes="synonym: b0076"
646..1767
/genes="leuO"
/function="putative regulator; Amino acid biosynthesis:
Leucine"
/notes="o373; 100 pct identical to LEUO_ECOLI SW:
P10151(290 aa) but contains 59 additional N-ter aa and 24
C-term residues; 100 pct identical to PIR: S40589 but
contains 24 additional C-term residues"
/codon_start=1
/transl_table=11
/product="probable transcriptional activator for leuABCD
operon"
/protein_id="AAC73187.1"
/db_xref="GI:1786264"
/translation="MPHSTAMDSVFIRTFEMSEFYSFCFLFYMHDKSYSGSLFLC
IPRRELSTVYLSMPVQTDHPETAELSKPOLRMVDNLNLTVDFAWQEQNITRAA
HVLGMSQPAVSNALVKVFNDELFVYRGIGQTPARAFQFGSVRQALQVONELP
GSGFEPASSERAVHLKVCPLDSILTSQIYNHIEQIAPNHYMFKSLNQNTQHOLRY
QETFFVISYEDHREFTSVPLFKDEMVLVASKNHPTIKGPLLKHVYNEQHAASLD
RFASFQPMYDITDKOASTAYOCGMMMSVLSVVSQTHLVALAPRWLAEEFAESLELQV
LPLKQNSRTCYLSHHEAAGRKDGHQWEEQLVLSICKR"
1802..1816
/notes="central position to predicted promoter: -39.5"
/bound_moiety="Lrp predicted site"
1813..1839
/notes="factor Sigma70; promoter ilvIHp4; documented+1 at
85394"
1833..1847
/notes="central position to predicted promoter: -8.5"
/bound_moiety="Lrp predicted site"
1838..1866
/notes="factor Sigma70; promoter ilvIHp3; documented+1 at
85420"
1915..1929
/notes="central position to predicted promoter:73.5"
/bound_moiety="Lrp predicted site"
1949..1963
/notes="central position to predicted promoter:107.5"
/bound_moiety="Lrp predicted site"
1955..1984
/notes="factor Sigma70; promoter ilvIHp2; documented+1 at
85534"
1978..1992
/notes="central position to predicted promoter:136.5"
/bound_moiety="Lrp predicted site"
1995..3809
/genes="ilvI"
/notes="synonym: b0077"
1995..3809
/genes="ilvI"
/EC_number="4.1.3.18"
/function="enzyme; Amino acid biosynthesis: Isoleucine,

```

```

Valine"
/notes="o604; 98 pct identical (1 gap) to 522 residues
from ILVI_ECOLI SW: P00893 (566 aa)"
/codon_start=1
/transl_table=11
/product="acetolactate synthase III, valine sensitive,
large subunit"
/protein_id="AAC73188.1"
/db_xref="GI:1786265"
/translation="MPWSLAANCLSKIGRLMCTHPRTQVROAMEMLSGAEVVRSL
IDGVKQVFGPGGAVLDIYDALHTVGGIDHVLVRHEQAHHADGMLARATGEVGVVL
VTSQPGATNAITGIATAYMDSIPLVLSQVATSLIGYDAFQECDMVGISRPVVKHSF
LVQETDIPQLKKAFLAASGRPGVVDLPKIDILNPANKLPYVMPESVMSRYNPT
TTGHQGIKRALQTLVAAKPPVYVGGGATAGCHQOKETVETALNLPVVCSLMGLGA
FPATHROALGMHGTYEANTMHNADVIFAVGVRFDRTTNNLAKYCNATVLIHD
IDPTSLSKVTADIPVGDAROVLEOMLELSOESAHOPLDEIRDHWQOLEORABQC
LYDTHSEKIKPOAVIETLWRLTKGDAYVTSYDQGHQMFALYIPDKPRWLNNSGL
GTMGFGLPALGVKMALEPVTVCVTGDSIQMNIQELSTALQYELPVLVNLNNRYL
GMVKQDMYISGRHSQSYMOSLPDFVRLAEAYGHVGIQISHPHELESKLEALEQVR
NNRLVFVDVTVDGSEHVYPMQIRGGGMDENWLSKTKERT"
1998..2012
/genes="ilvI"
/notes="central position to predicted promoter:156.5"
/bound_moiety="Lrp predicted site"
2017..2045
/genes="ilvI"
/notes="factor Sigma70; promoter ilvIHp1; documented+1 at
85597"
3812..4303
/genes="ilvH"
/notes="synonym: b0078"
3812..4303
/genes="ilvH"
/EC_number="4.1.3.18"

Query Match      19.4%; Score 282; DB 1; Length 21757;
Best Local Similarity 53.2%; Pred. No. 6.8e-25;
Matches 731; Conservative 0; Mismatches 610; Indels 33; Gaps 5;

QY  80  CATCGTTGTGCGGCTCGGCAAGAGCGGCTGTCCTGGTGGCTACCTGGCGGCGCGG 139
DB  13568  CGTCATATCGGCTGGGCTCACCGGGCTTTCTCGGTGGAGATTTTCTCGCGG 13627

QY  140  CTTGCCCTTTCGCGGTGCGATACCCGAGAGAACCCGCGGAGCTGCCACCTCGGTGC 199
DB  13628  TGTGAGCGCGCGGCTTATGATACCGGTATGACACCGCTGCGCTGATAATATACCG- 13686

QY  200  CCAATATCCGACGTTGGAAGTGGCTTTCGCGGGAACCTGACGCGCGAGTTCCTCTGCTCCG 259
DB  13687  -----AAGCGGTAGAAGCCACACGCGGCACTCTGAATGATGATGGCTGATGGCGGC 13738

QY  260  CCGCGAATCTATGTGTCAGCCCGGCTTGTGCTGCGGCAACCCCTGCGCTGTACAGGCGCG 319
DB  13739  AGATCTGATGTGCGCCAGTCCCGGTATGACACTGGCGCATCATCTTAAGCGCTGCCG 13798

QY  320  CGCGAAAGCGTGGCATCTCCGGTGACATGATCTTTGCGCCCGGAGCGAGGCGCC 379
DB  13799  TGATGCCGAATCGAAATCGTTGGCATATCGAGCTGTTCTGTCGCGAAGCACAAGCACC 13858

QY  380  GATCGTCGCCATCACCAGTTCCAAACCGGAGACCGGTGACCAACCTGTGGCGGCAAT 439
DB  13859  GATTGTGGGATTTACCGGTTCTAACGGCAAAAGCAGCGGTACCAACCTAGTGGGTGAAAT 13918

QY  440  GCGCGTGGCGGCAAGCGTGTGCGCGGCAACCTGCGGCAACCTGCCCGCGCTCGA 499
DB  13919  GCGGAAAGCGCGGGGGTTAACGTTGTTGGGTGGCAATATTGGGCTGCTCGCTGAT 13978

QY  500  CCTGCTGGCGGACGACATCGAGCTGACGTGTTGGAGCTGTTCGAGCTTCAGCTGGAAC 559
DB  13979  GCTACTGGATGATGAGTGTGAACGTGACGTGCTGGAACGTGTCGAGCTTCAGCTGGAAC 14038

QY  560  CTGCGATCCCTCAACGCCGAGGTGGCGACCGTGTGACGCTGACGCGAGACCATATGGA 619
DB  13979  GCTACTGGATGATGAGTGTGAACGTGACGTGCTGGAACGTGTCGAGCTTCAGCTGGAAC 14038

```

Db 14039 CACCTCCAGCTTACAGCGGTAGCAGCGACCAATCTTGAACGTGACTGAAGATCATATGGA 14098  
Qy 620 TCGCTA---CGACGSCATGGCTGACTACCACTGGCCAGCAGCGATCTCCGCGTGC 676  
Db 14099 TCGCTATCCGTTGGTTTACAAAGATATGTCGAGCAAAATGCGCATTTACGAAACGC 14158  
Qy 677 CCGCAGGTGCTGTGATGATCGCGCGATGCCCTGACCGAGCGCTGATCGCGCATACCGT 736  
Db 14159 GAAAGTTTGGTGTATGCTGATGATGCTTAAACAATCCGATTCGCGTGGGATGA 14218  
Qy 737 GCCGTGCTGGTTCGCGCTGAAACAGCGGAGCTTCAAGGCTTTCGGCTGATFCGAGGA 796  
Db 14219 ACGTTCGGCTGAGCTTTGGCGTCAACATGGGTGACTATCA-----CCTGAATCATCA 14269  
Qy 797 AGACGGCAGAGTGGCTGGCGTTCAGTTCGACAACTGCTGCCGTTGGCGAAGTAA 856  
Db 14270 GCAGGGCAACCTTGGCTGGGGTTAAGCGGAGAAAGTCTGAATGTGAAGAGATGAA 14329  
Qy 857 GATCCGTGGGCGCCACAACTATTCCAAAGCGCTGCGCGCTGGCGCTGGGCCATGCGGT 916  
Db 14330 ACTTTCGGCGAGCATAACTACACCAATGCTGGCGCGCTGGCGCTGGCAGATGCTG 14389  
Qy 917 CGGCTCGCGTTCGACGCCATGCTCGCGCGCTGAAAGCGTTCCTGCGCTGCGCTCATCG 976  
Db 14390 AGGTTTACCGCTGCCAGCGCTGAAAGCGTTAAACCACTTCACTGCTGCGCGCATCG 14449  
Qy 977 CTGCACTGGGTACGCGAGCGGCGAGGCGCTGAGCTACTACGACGATTCGAAGGCCACAA 1036  
Db 14450 CTTTGAAGTTGCTGAGGATACAGCGGTACGTTGATTAACGATTCGAAGGCCACAA 14509  
Qy 1037 CGTGGCGCGCGCTGCGCGGATFCGAGGGGCTGGTTCGCGACATCGACGCGCAAGCTGGT 1096  
Db 14510 CGTGGCGAGTACGGAAGCGCGCTGAATGGCGTG-----CACGTAGACGCGACACTGCA 14563  
Qy 1097 GCTGCTCGCGCGGAGACGCGAGCGCGCGATTTCCATGACCTGCGCGAGCGCGTGC 1156  
Db 14564 TTTGTTGCTGGGTGGCGATGATAAATCGCGGAGCTTTAGCCCACTGCGCGGTTACCTGAA 14623  
Qy 1157 GCGCTTTCGCGCGGTGGTACTGCTTGGCGGTGACGCGCGGTGATTCGCCAGGCACT 1216  
Db 14624 TGGGATAACGTAGCTGTATTGTTTCGTTGCTGAGCGCGCGAGCTGGCGGCTAC- 14682  
Qy 1217 GGGCAACGGGTACCGTGGTTCGCGCTGCGCAACGCTGGAGCAAGCATTCGCGGAGCGCG 1276  
Db 14683 -----GCCCGGAGTGGCAGAAACAAACCGAACTATGGAACAGCGCATGGCGTTGCTGGC 14737  
Qy 1277 CGAGCTGCGCGGAGGCGATGGGTGCTGTTGCTGCGCGCGCTGCGCGAGCTGGACAT 1336  
Db 14738 TCCCGGTTCAGCGGGCGATATGGTCTGCTCTCCCGAGCTGTGCCAGCTTGTATCA 14797  
Qy 1337 GTTCAAGAACTTCGAAGACGCGGACGCTGTTTCCCAAGCCGTAGAGAGCTAGCGTG 1396  
Db 14798 GTTCAAGAACTTTGAACACAGGCAATGATTTGCCGCTGCGGAGGATGAGTTG 14857  
Qy 1397 ATGCTGCTGGTTCGCGCGCTTCCCGTGGCGCTGTTGAGCGCGGACGGCATC 1450  
Db 14858 ATGCGTTATCTCTCCCTCGCCTGAAATGCGCGCGCTGCGAGGATTCAGTATC 14911

RESULT 15  
AX191720 AX191720 28277 bp DNA linear PAT 15-AUG-2001  
DEFINITION Sequence 2 from Patent WO0149775.  
ACCESSION AX191720  
VERSION AX191720.1 GI:15209889  
KEYWORDS  
SOURCE Escherichia coli  
ORGANISM Escherichia coli  
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
Enterobacteriaceae; Escherichia.  
REFERENCE  
1 Iversen, P.L.  
AUTHORS Antisense antibacterial cell division composition and method  
TITLE

JOURNAL Patent: WO 0149775-A 2 12-JUL-2001;  
Avi Biopharma, Inc. (US)  
FEATURES Location/Qualifiers  
source 1..28277  
/organism="Escherichia coli"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:562"  
BASE COUNT 6714 a 6725 c 7993 g 6845 t  
ORIGIN  
Query Match 19.4%; Score 282; DB 6; Length 28277;  
Best Local Similarity 53.2%; Pred. No. 6.2e-25;  
Matches 731; Conservative 0; Mismatches 610; Indels 33; Gaps 5;  
Qy 80 CATCGTTGTCGGCCCTCGCAGAGCGGCATGCTCCCTGGTGGCTACCTGGCGCGCGCGG 139  
Db 13643 CGTCATTATCGCTGGCGCTCACCGGCTTCTCTGGTGACATTTTCTCTGGCTCGCGG 13702  
Qy 140 CTTGCCCTTTCGGCTGCTCGATACCCGAGAGAACCCGCGGAGCTGGCCACCCCTCGCTGC 199  
Db 13703 TGTGACCGCGCGCTTATGGATACGCTATGACACCGCTGCTGGCTGGATAAATACCG- 13761  
Qy 200 CCAATATCCGAGGTGAAGTGGCTTGGGGGAACTGACGCCGAGTTCCTCTGCTCGCG 259  
Db 13762 -----AAGCGCTAGAACGCCACACGGGCGAGTCTGAATGATGAATGGCTGATGGCGC 13813  
Qy 260 CCGGAACTCTATGTCAGCCCGCTTGTGCTGCGCACCCCTGCTGGTGTACAGCGCGC 319  
Db 13814 AGATCTGATGTCGCCAGTCCCGGTATGCACTGGCGTCCATCTTAAAGCGCTGCGC 13873  
Qy 320 CGCGAAAGGCTGCGCATCTCCGGTGACATGATCTCTTCGCCGCGAGGCGAAGGCCCC 379  
Db 13874 TGATGCGGAATCGAAATCGTTGGCATATGAGCTGTTCTGTCGCGAAGCACAGCAC 13933  
Qy 380 GATCGTCGCCATCACCGGTTCCACGCGAAGACACCGTACCACCTGGTGGCGGAAT 439  
Db 13934 GATTTGCGGATTTACCGGTTCTTAACGCGAAGACACCGTACCACGCTAGTGGGTGAA 13993  
Qy 440 GCGCGTGGCGCGGACAGCGTGTGCGCTGCGCGGCAACCTCGCACCCCGCGCTCGA 499  
Db 13994 GCGAAAGCGCGGGGTTACGTTGGTGGGTGGCAATATTGGCTGCTCGCTTGAT 14053  
Qy 500 CCGTCTGCGCGACACATCGAGCTGTAGTGTGGAGCTGTGAGCTTTCAGCTTTCAGTGA 559  
Db 14054 GCTACTGATGATGAGTGAATGATGCTGCTGAACTGCTGAACTTCCAGCTTCCAGCT 14113  
Qy 560 CTGCGATCGCTCAACCGCGAGTGGCGACCGTGTGACGCTGACGAGACCATATGGA 619  
Db 14114 CACCTCCAGCTTACAGCGGTAGCAGCGACCATTTCTGAACGTGACTGAAGATCATATGA 14173  
Qy 620 TCGCTA---CGACGCGATGGCTGACTACACCTGGCCAGCAGCGGATCTTCCGCGTGC 676  
Db 14174 TCGCTATCCGTTTGGTTTACAAAGTATCGTGCAGCAAACTGCGCATTTACGAAACGC 14233  
Qy 677 CCGCCAGTGTGTGATTCGCGCGGATGCGCTGACCGAGCGCTGATFCGCGGATACCGT 736  
Db 14234 GAAAGTTTGGTGTGATGATGCTGATGCTTAACTAACAATGCGGATTCGCGTGGGATGA 14293  
Qy 737 GCGCTGCTGGTTCGCGCTGAAACAGCGCGGCTTCAAGGCTTTCGCGCTGATCGAGGA 796  
Db 14294 ACGCTGCTGAGCTTTGGGTCACATGGGTGACTATCA-----CCTGAATCATCA 14344  
Qy 797 AGACGGCAGAAAGTGGCTGGGTTCCAGTTTCGCAAGCTGCTGGCGTTGGCGAAGTAA 856  
Db 14345 GCAGGGCAACCTTGGCTGGCGGTTAAGCGGAGAAAGTGTGAATGTGAAGAGATGAA 14404  
Qy 857 GATCCGTGGCGCCACAACTATTCCAAAGCGCTGCGCGCTGGCGCTGGCGGCTGCGGT 916  
Db 14405 ACTTTCGGGCGAGCATAACTACACCAATGCTGCGCGCGCTGGCGCTGGCAGATGCTGC 14464  
Qy 917 CGGCTGCGCTTCAGCGCCATGCTGCGCGCGCTGAAGCGGCTTTCCGCGCTGGCTCATCG 976  
Db 14465 AGGTTTACCGCTGCCAGCGCTGAAAGCGGTTAAACCATTCATCTGCTGCTGCGCGATCG 14524

```

Qy 977 CTGCCAGTGGGTACGCGAGCGGAGGCGGTGAGCTACTACGACGATTCCAAAGGCCACCAA 1036
Db 14525 CTTTGAAGTTGTGCTGAGCATTAACGCGTACGTTGGATTACGATTTCGAAAGCGACCAA 14584
Qy 1037 CGTCGGCGCGCCTGCGCGGATCGAGGGCTGGGTGCCGACATCGACGGCAAGCTGGT 1096
Db 14585 CGTCGGCGAGTACGGAAGCGCGCTGAATGCGCTG-----CAGGTAGACGGCACACTGCA 14638
Qy 1097 GCTGCTCGCGCGGAGAGCGGCAAGGCGCGGATTTCCATGACCTGCGCGAGCGGTCGC 1156
Db 14639 TTTGTTGCTGGGTGGCGATGGTAATCGGCGGACTTTAGCCCACTGGCGGTTACCTGAA 14698
Qy 1157 GCCTTCTGCGCGGCGGTGCTACTGCTTGCCCGTGAAGCGCGGCTGATTTGCCCAGGCACT 1216
Db 14699 TGCGGATAACGTACGTCTGTATTGTTTCGGTCTGTGACGGCGCGCAGCTGGCGGCTAC- 14757
Qy 1217 GGGCAACGCGGTACCGCTGTGCGGCTGCGCAACGCTGGAGCAAGCAGTCCGGCAGGCCGC 1276
Db 14758 -----GCCCGGAAGTGGCAGAACAAACCGAAACTATGGAAACAGGCGCATGCGTTGCTGGC 14812
Qy 1277 CGAGCTGGCGCGGAGGCGGATGCGGTGCTGTTGTCGCGGCGCTGCGGAGCCTGGACAT 1336
Db 14813 TCCGCGTGTTCAGCGGCGGATATGTTCTGCTCTCCCCAGCCTGTGCCAGCCTTGATCA 14872
Qy 1337 GTTCAAGAACTTCGAAGAACGCGAGCGCCTGTTGCGCAAGCGGTAGAGAGCTAGCGTG 1396
Db 14873 GTTCAAGAACTTGAACAGCAGGCAATGAGTTTGCCCGTCTGGCGAAGGAGTTAGGTTG 14932
Qy 1397 ATGCTGTCGGTGTGCGCCCTTCCCGTCCCGCTGTTGAGCCGCGCACGGCATC 1450
Db 14933 ATGCGTTTATCTCTCCCTGCGCTGAAATGCCCGCGCTGCCAGGATTTCAGTATC 14986

```

Search completed: August 14, 2003, 03:24:00  
Job time : 5318 secs